

Calibration of roughness for flood modeling using satellite data
Caractérisation des rugosités par données satellitaires pour les calculs d'inondations

École doctorale N°531, Sciences, Ingénierie et Environnement (SIE)

Spécialité du doctorant : Sciences de l'ingénieur

Thèse préparée au sein du Laboratoire d'Hydraulique Saint-Venant
(École nationale des ponts et chaussées et EDF R&D)

Thèse soutenue le 14/10/2025, par
Jean-Paul TRAVERT

Composition du jury:

Eric BLAYO
Professeur, Inria & UGA, LJK

Rapporteur

Renaud HOSTACHE
Directeur de Recherche, IRD, UMR Espace-Dev

Rapporteur

Florence TUPIN
Professeure, Télécom Paris, LTCI, IP Paris

Présidente du jury

Hélène ROUX
Professeure, Toulouse INP, IMFT

Examinatrice

Sébastien BOYAVAL
ICPEF, LHSV, ENPC, IP Paris & Inria

Directeur de thèse

Cédric GOEURY
Ingénieur Chercheur, EDF R&D & LHSV, ENPC, IP Paris

Co-Encadrant de thèse

Vito BACCHI
Ingénieur Chercheur, EDF R&D

Co-Encadrant de thèse

Fabrice ZAOUÏ
Ingénieur Chercheur, EDF R&D

Co-Encadrant de thèse

Philippe Moireau
Professeur, CMAP, Ecole Polytechnique, IP Paris & Inria

Invité



Thèse effectuée au sein du Laboratoire d'Hydraulique
Saint-Venant de l'École nationale des ponts et chaussées et
d'EDF R&D, 6 quai Watier, BP 49, 78401 Chatou cedex,
France.



INSTITUT
POLYTECHNIQUE
DE PARIS



Convention CIFRE 2022/0972

Remerciements

Cette thèse représente trois ans de travail personnel, mais aussi trois ans de rencontre, de collaboration et de soutien. Sans vous tous, cette aventure n'aurait jamais eu la même saveur.

Tout d'abord, ma première pensée se dirige vers mon directeur de thèse Sébastien ainsi que mes encadrants Cédric, Vito et Fabrice. À votre contact, j'ai pu grandir scientifiquement, mais aussi humainement. Au-delà d'un simple encadrement, je vous considère comme une équipe unie dans laquelle je me suis épanoui pendant trois ans. Vous avez toujours eu beaucoup de bienveillance et vous m'avez toujours encouragé à poursuivre mes idées et à ne pas me brider. Je vous remercie encore pour votre disponibilité et l'ensemble des bons moments passés à vos côtés. Nous pouvons être fiers de cette petite équipe et du résultat de ce travail mené ensemble, même si la tradition veut que seul mon nom figure sur ce manuscrit ! Quand je repenserai à ma thèse dans quelques années, je ne me souviendrai peut-être pas parfaitement des méthodes de traitement des images satellitaires ou de certaines subtilités sur l'assimilation de données. Mais, il est certain, que je me souviendrai de nos moments face aux équations sur les tableaux blancs, de votre enthousiasme à chaque réunion, des digressions plus ou moins scientifiques. J'espère que nous aurons encore l'occasion de collaborer dans les prochaines années.

Ensuite, je souhaite remercier mon comité de suivi de thèse, Pierre-Olivier Malaterre et Alban Farchi pour avoir suivi mon travail avec beaucoup d'intérêt. Je souhaite également remercier sincèrement Eric Blayo et Renaud Hostache pour avoir accepté d'être rapporteur et pour leur relecture attentive de mon manuscrit. Vos questions et vos remarques ont aidé à améliorer mon travail et de lui donner une portée encore plus significative. À Florence Tupin, Hélène Roux et Philippe Moireau, merci pour ces échanges passionnants lors de la soutenance : vos questions m'ont permis de prendre du recul et d'élargir mes perspectives. Ce jury d'exception restera gravé dans ma mémoire, tant pour son exigence que pour sa bienveillance et son enthousiasme.

Une thèse, c'est aussi l'ouverture vers d'autres horizons. Merci à Sylvain Biancamaria pour son suivi et ses conseils sur l'imagerie satellitaire, ainsi qu'aux équipes du CNES (Lionel Zawadzki, Nicolas Gasnier, Santiago Pena Luque) et du CERFACS (Ludovic Cassan) pour nos échanges stimulants sur cette thématique. Je garde également un excellent souvenir des discussions avec Philippe Moireau, Damiano Lombardi et Gaël Le Ruz sur des nouvelles pistes concernant l'assimilation de données, un axe de recherche, bien que difficile et connexe à mes travaux, mais sur lequel j'ai trouvé beaucoup de plaisir à travailler dans mes derniers mois de thèse. Merci à Urbain Vaes pour nos échanges passionnants sur l'assimilation de données, et à Anju Sebastian et Jérôme Thiébôt pour m'avoir ouvert la porte à une thématique nouvelle : l'estimation de la ressource hydrolenne. Ces collaborations et discussions ont été autant de bouffées d'air frais dans mon travail.

Ce travail s'inscrit également au sein d'une équipe plus large, d'abord le projet MOISE (modélisation des inondations externes) et plus globalement l'ensemble des collègues des groupes P73 et P74 au sein du Laboratoire National d'Hydraulique et Environnement (LNHE) et de EDF R&D. Merci pour les échanges techniques, mais aussi pour les pauses

café, les déjeuners et les moments conviviaux qui permettent de traverser la thèse plus simplement. Une pensée particulière pour Vito et Florent, qui ont cru en moi dès mon premier entretien de stage et m'ont permis de vivre cette belle aventure. Merci également au Laboratoire d'Hydraulique Saint-Venant (LHSV) pour son accueil chaleureux et ses événements scientifiques enrichissants.

À mes compagnons de route, merci pour votre présence et votre bonne humeur : mon super camarade de promo, Guillaume ; les doctorants, plus ou moins jeunes, Elisa, Bastien, Fatna, Antoine, Prokopios, Jiaze, Mohammed, Alessandro, Faten, Claudine, Zied, Ahmed, Arthur, Yohan, Jiankai, Jules et Clément ; ainsi que l'équipe éoliennes, Maxime, Titouan et Fabien. J'ai également une pensée pour l'ensemble des stagiaires et co-bureaux qui ont apporté renouveau, dynamisme et bonne humeur tout au long de cette aventure : Morgane, Anh Ly, Sébastien, Dimitri, Azélis, Claudine, Lucas, Benjamin, Guillaume, Jade, Paul-Édouard, Camille, Kévis, Valentin, Clément et Kosuke. J'ai sûrement, et malheureusement, oublié quelques personnes dans cette liste, mais merci à toutes et tous pour les moments agréables partagés en votre compagnie. Merci pour les discussions, les fous rires, les parties de frisbee, les activités sportives, les week-ends, les afterworks... Ces souvenirs resteront précieux.

Enfin, cette aventure n'aurait pas eu la même saveur sans mes amis avec lesquels j'ai pu partager de nombreux moments pendant ces trois années. Je pense en particulier à mes amis de l'ENSE3, Quentin, Alex, Léo, Jeanne, Estelle, Thomas, Augustin, Guillaume, mais aussi plus largement autour de ce petit groupe à Maëlle, Alexandra, Laura, Julie, Jérémy, Célia et Félix. Merci pour votre amitié et votre soutien indéfectible. Merci également aux cigales (Paul, Laurène et Loïse) pour nos week-ends annuels, les discords, et l'ensemble des bons moments depuis notre rencontre en 2020. Et bien que nous avons eu moins l'occasion de se voir ces trois dernières années, je pense à mes amis de la région valognaise et cherbourgeoise (Thomas, Roxane, Romain, Louise, Pierre, Lucas, Anaïs, Guillaume, Sophie et Baptiste) avec qui j'ai toujours un grand plaisir à me retrouver.

Et bien sûr, à mes parents et à ma sœur, merci d'avoir toujours cru en moi et de m'avoir donné les moyens de poursuivre mes rêves. Cette réussite est aussi la vôtre.

Résumé

La simulation des inondations repose souvent sur des modèles hydrodynamiques bidimensionnels (2D), qui doivent être soigneusement calibrés et validés à l'aide d'événements d'inondation passés afin de garantir la précision des prévisions futures. Le coefficient de rugosité est un élément essentiel de ce processus de calibration, car il influe considérablement sur les profondeurs d'eau et les vitesses dans le modèle. Dans le lit mineur, les modèles hydrodynamiques sont calibrés à l'aide de données chronologiques provenant de stations de jaugeage. Mais pour les plaines d'inondation, la distribution spatiale du coefficient de rugosité est souvent mal connue et généralement déduite à partir de l'occupation des sols, associée à des valeurs empiriques tirées du jugement d'experts. La calibration des modèles hydrodynamiques pourrait bénéficier de l'intégration du maximum de données d'observation disponibles.

Dans ce contexte, la télédétection par satellite peut être exploitée comme une source précieuse d'informations. L'imagerie par radar à synthèse d'ouverture (SAR), en particulier, est capable de détecter les étendues d'eau de surface même sous la couverture nuageuse, ce qui la rend très adaptée à la surveillance des inondations. Cette thèse vise à développer une méthodologie pour calibrer les coefficients de rugosité des plaines d'inondation dans les modèles hydrodynamiques 2D, en tirant parti des observations dérivées des satellites.

La première étape vise à récupérer des informations à partir des observations satellitaires pour calibrer les modèles d'inondation. Ainsi, cette thèse propose une évaluation de l'extraction des cartes d'étendue des inondations et des champs de profondeur d'eau à partir de l'imagerie SAR, ainsi qu'une évaluation approfondie des incertitudes associées, découlant des choix de méthodes et de leurs hyperparamètres. Les cartes d'inondation observées doivent être comparées au modèle hydrodynamique afin de le calibrer. Nous avons développé une méthodologie pour guider la sélection d'une mesure de performance appropriée afin de comparer les étendues d'inondation simulées et observées. Finalement, les observations incertaines dérivées des images satellites SAR sont combinées avec le modèle hydrodynamique grâce à des techniques d'assimilation de données, afin d'améliorer la calibration des coefficients de rugosité variables dans l'espace sur les plaines d'inondation.

En combinant les données de télédétection, la quantification des incertitudes et la modélisation hydrodynamique, cette recherche fournit une approche pour améliorer la calibration des modèles d'inondation, avec le potentiel de renforcer la prévision des inondations et les pratiques de gestion des risques.

Mots-clés : Modélisation des inondations, Observations satellitaires, Assimilation de données, Quantification d'incertitude, Mesures de performance, Rugosité

Abstract

Flood simulation commonly relies on two-dimensional (2D) hydrodynamic models, which must be carefully calibrated and validated using past flood events to ensure accurate forecasting. A critical component of this calibration process is the spatial distribution of the roughness parameter, which significantly influences simulated water depths and velocities. In the minor bed, the hydrodynamic models are calibrated using time-series data from stage-gauging stations. For floodplains, few data are available, and the roughness spatial distribution is often poorly known and typically inferred from land-use classifications using empirical values drawn from expert judgment. Hydrodynamic models' calibration could benefit from incorporating the maximum available observational data.

In this context, satellite remote sensing can be leveraged as a valuable source of information. Synthetic Aperture Radar (SAR) imagery, in particular, is capable of detecting surface water extents even under cloud cover, making it highly suitable for flood monitoring. This thesis aims to develop a methodology to calibrate floodplain roughness coefficients in 2D hydrodynamic models, leveraging satellite-derived observations.

The first step aims to recover information from the satellite observations to calibrate the flood models. Thus, this thesis proposed an evaluation of the extraction of flood extent maps and water depth fields from SAR imagery, along with an evaluation of the associated uncertainties arising from method choices and hyperparameter settings. The flood extent maps should be compared with the hydrodynamic model to calibrate it. We developed a step-by-step methodology to guide the selection of appropriate performance measures to compare simulated and observed flood extent maps. Then, the uncertain SAR-derived flood observations are integrated into the hydrodynamic modeling framework through data assimilation techniques to enhance the calibration of spatially distributed roughness coefficients on floodplains.

By combining remote sensing data, uncertainty quantification, and hydrodynamic modeling, this thesis proposes an approach to enhance flood model calibration, with the potential to improve flood forecasting and risk management practices.

Keywords: Flood modeling, Satellite observations, Data Assimilation, Uncertainty Quantification, Performance measures, Roughness

Contents

1	Introduction	14
1.1	General context	14
1.2	Industrial context	16
1.3	Scope of the thesis	17
1.4	Outline of the manuscript	18
1.5	Scientific productions	19
2	Flood Modeling and Satellite Observations	21
2.1	Modeling free-surface flows	23
2.2	Monitoring floods with satellite observations	37
2.3	Summary	59
3	Uncertainty Quantification and Data Assimilation	61
3.1	General notations	64
3.2	Uncertainty Quantification	65
3.3	Introduction to Data Assimilation	74
3.4	Bayesian approach	77
3.5	Bayesian approach for time-dependent models	79
3.6	Variational approaches	85
3.7	Summary	92
4	Flood extent and water depth from SAR observations	93
4.1	Motivations	94
4.2	Preprint under revision	94
4.3	Additional Case Study	125
4.4	Graphical User Interface	130
4.5	Summary	131
5	Evaluation of Performance Measures for Comparing Flood Models with Satellite Observations	132
5.1	Motivations	133
5.2	Published article in Water Resources Research	133
6	Calibration of roughness field in 2D flood model with flood maps assimilation	174
6.1	Motivations	175
6.2	Preprint in preparation	175
6.3	Complements: Twin experiment on a schematic test case	203
6.4	Summary	203

7	Calibration of roughness field in 2D flood models with water depth fields assimilation	204
7.1	Motivations	206
7.2	Water depth assimilation on a schematic test case	206
7.3	Water depth assimilation on Garonne River	217
7.4	Summary	226
8	Conclusions and Perspectives	227
8.1	Conclusions	227
8.2	Perspectives	231
	List of Acronyms	235
A	Appendix	238
A.1	Complements to Chapter 6 on a schematic test case	238
A.2	Data assimilation with Optimal Transport on manifolds	247

List of Figures

1.1	Water depth on a two-dimensional simulation with two spatial configurations of the roughness parameter.	15
1.2	(a) Optical satellite image (MODIS mission) in Bangladesh before flood on June 5 (left) and during flood on July 25 (right) 2022, (b) Radar satellite image (Sentinel-1 mission) before flood on January 16 (left) and during on February 3 (right) 2021 over the Garonne River.	16
1.3	Aerial view of the Blayais nuclear power plant with its four production units and the protective dikes. <i>Credit: Didier Marc.</i>	17
2.1	Schematic of the free surface, bottom, and shear stresses.	25
2.2	openTELEMAC architecture adapted from Goeury et al. [2022].	28
2.3	Examples of photographs of stream channels with associated estimated Manning (n) from the Manning-Strickler formula. Credit: U.S. Geological Survey.	31
2.4	Corine Land Cover classification of land use in France.	34
2.5	Satellite imagery of various natural hazards: (a) Wildfires in Albania (28 July 2025); (b) Typhoon Khanun impacting Japan (3 August 2023); (c) Vegetation resurgence after rainfall in Morocco (27 March 2025); (d) Skeiðarárjökull glacier, Iceland (9 July 2025); (e) Mount Etna eruption (2 June 2025); (f) Flooding after heavy rainfall in Pas-de-Calais, France (12 November 2023). Credit: European Union, Copernicus Sentinel-2 imagery.	38
2.6	Difference between passive (left) and active (right) satellite sensors.	39
2.7	Electromagnetic spectrum with associated wavelengths (top arrow) and frequencies (bottom arrow).	39
2.8	Comparison of a Synthetic Aperture Radar Sentinel-1 image (left) with an optical Sentinel-2 image (right) acquired at the same time.	41
2.9	Acquisition geometry of a side-looking radar imager. The radar moves along the azimuth direction and emits pulses in the Line Of Sight direction.	43
2.10	Principle of a Synthetic Aperture Radar sensor capturing the same point P multiple times with an antenna of length L_A to create a “synthetic” antenna using the Doppler effect. The shades of grey show the number of times the same area is seen in the swath.	45
2.11	Illustration of speckle noise on a Sentinel-1 image on the left and without speckle on the right. The black zones correspond to water, and the brighter zones to pasture areas.	46
2.12	Terrain distortions due to the incidence angle and topography.	48
2.13	Scattering mechanisms of Synthetic Aperture Radar signals on various types of targets.	48
2.14	Polarization of the electric field of a planar wave.	50

2.15	Influence of local height variations, incidence angle, wavelength, and canopy density on backscatter mechanisms adapted from Lillesand et al. [2015] and Landuyt [2021].	51
2.16	Mind map of flood mapping with Synthetic Aperture Radar observations. .	52
2.17	Overlay of a flood map (red contour) on a Digital Elevation Model with contour lines every 0.5 m of elevation.	57
2.18	Example of schematic cross-section analysis with a Digital Elevation Model and a flood map (red contour) to estimate the water depth.	58
3.1	Sketch representing the different steps in an Uncertainty Quantification study adapted from Baudin et al. [2015].	67
3.2	Example of Design of Experiment with Random Sampling, Latin Hypercube Sampling, and low discrepancy sequence in the search space $[-1, 1]^2$ with uniform input distributions.	68
3.3	Example of $r = 4$ trajectories in two dimensions on the Hypercube $[0, 1]^2$ with $L=9$ levels and a discretization step of $1/8$. The filled circles represent the start of the trajectories.	70
3.4	Principle of Data Assimilation and notations of the problem on a schematic example of a flood event.	75
3.5	Sequential Data Assimilation principle with a Kalman Filter. The analyzed state $\mathbf{x}_k^a$ is corrected with $\mathbf{x}_k^f$ and observation $\mathbf{y}_k^{obs}$. The forecast $\mathbf{x}_k^f$ comes from model propagation starting from $\mathbf{x}_{k-1}^a$	79
3.6	Sequential Data Assimilation with Ensemble Kalman Filter. Analysis samples are computed (blue crosses) based on the observation $\mathbf{y}^{obs}$ (orange hexagons) and the forecasted ensemble $\mathbf{x}^f$ (green dots). At each time step, the analysis ensemble is propagated by the model (green trajectories). . . .	82
3.7	Three-Dimensional Variational (3D-Var) and 4D-Var assimilation scheme representation. The aim is to determine the optimal analysis trajectory (dashed blue curve) from a background trajectory (green curve) and a set of observations (orange marks).	86
4.1	Workflow for processing Sentinel-1 (SAR) images to derive flood maps. Raw images are preprocessed using 26 hyperparameter configurations (Table 2 in Section 4.2), then each image is turned into 48 flood maps (or 432 if using morphological operations) (Table 3 in Section 4.2).	125
4.2	Study area and flood map produced by the Copernicus Emergency Mapping Service on the study area using a COSMO-SkyMed image acquired on January 9, 2016, at 18:39 UTC.	126
4.3	Accuracy and F1-score computed between the generated flood maps and the Emergency Mapping Service reference map grouped by flood mapping method, along with the flooded surface area for the generated flood maps. .	127
4.4	Range of variation (between the minimum and maximum values) of F1-score (a) and flooded area (b) due to filtering methods hyperparameters for the 48 flood mapping configurations. The points are jittered horizontally for visualization purposes.	128
4.5	Range of variation (between the minimum and maximum values) of F1-score (a) and flooded area (b) due to flood mapping methods hyperparameters for the 26 speckle filtering configurations. The points are jittered horizontally for visualization purposes.	129
4.6	Visualization of the Graphical User Interface for Sentinel-1 data downloading (a) and satellite image visualization (b).	131

4.7	Visualization of the Graphical User Interface for the preprocessing step. . .	131
6.1	Illustration of the schematic test case for preliminary tests of the methodology presented in Section 6.2.	203
7.1	Selected water depth observation locations for configuration 1 on the Garonne River: (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.	217
7.2	Selected water depth locations for the second configuration for water depth assimilation.	218
7.3	Generated water depth fields with FLEXTH method on the Garonne River: (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.	219
7.4	Total-order Sobol indices computed on the 92 floodplain subdomains for the water depth simulated at one location for (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.	219
7.5	Subdomains with total-order Sobol indices higher than 0.05 for (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.	220
7.6	Mean estimates of the Strickler values over the 10 initializations in the five subdomains over iterations of the 3D-Var, for (a) configuration 1 with 100 observation locations and (b) configuration 2 with 20 observation locations. The error bars represent the standard deviation of the estimates on the 10 runs.	221
7.7	Evolution of the mean Strickler values (over the 10 runs) over 3D-Var iterations for configuration 1, using 100 observation locations. Results are shown for varying observation errors across subdomains 0 to 3.	222
7.8	Evolution of the mean Strickler values (over the 10 runs) over 3D-Var iterations for configuration 2, using 20 observation locations. Results are shown for varying observation errors across subdomains 0 to 3.	223
7.9	Estimated Strickler values over 3D-Var iterations in subdomains 0 to 3, using water depth observations from satellite imagery. Assimilation was performed with noise levels of $\sigma^{obs} \in \{1.0\text{ m}; 1.1\text{ m}; 1.2\text{ m}\}$ for configuration 1 and $\sigma^{obs} \in \{0.5\text{ m}; 0.6\text{ m}; 0.7\text{ m}\}$ for configuration 2.	224
7.10	Water depth differences between the assimilated and reference simulations on February 2, 2021, at 17:56 UTC for (a) configuration 1 and (b) configuration 2.	225
7.11	Water depth differences between the assimilated and reference simulations on February 3, 2021, at 17:48 UTC for (a) configuration 1 and (b) configuration 2.	226
8.1	Example of a Large Scale Particle Image Velocimetry measure with FlowPic smartphone application (extracted from [Bodart, 2023]).	232
8.2	Differences between simulated and observed fire fronts (adapted from Rochoux et al. [2014]).	233
8.3	Example of intermediate states of water depth distribution on a cross-section with (a) Euclidean distance point to point, and (b) desired behavior.	234
8.4	Example of trajectory between a source distribution (left) and target distribution (right) using flow matching.	234
A.1	Input discharge and water depths simulated by the 1D model for two Strickler configurations.	250

A.2	Illustration of the Monge problem between two continuous distributions of same masses.	250
A.3	Monge transport between two discrete distributions ρ_0 and ρ_1	251
A.4	Monge-Kantorovich transport plan with mass splitting between two discrete distributions ρ_0 and ρ_1	252
A.5	(a) Transport map and (b) geodesics between two water depth distributions computed with a one-dimensional shallow-water code.	254
A.6	Verification of parallel transport implementation.	255

List of Tables

2.1	Proposed Manning values for river channels.	31
2.2	Proposed Manning values for floodplains.	32
2.3	Estimates of Manning’s n for five rivers with four estimation methods from Wohl [1998].	33
2.4	Proposed values of Strickler coefficient (K_s in $m^{1/3}s^{-1}$) for 1D and 2D models in the floodplains [Bacchi and Pavan, 2025].	35
2.5	Examples of widely used optical satellites for flood monitoring.	40
2.6	Examples of widely used altimeters for water level measurements.	41
2.7	Examples of widely used SAR satellites for flood applications.	42
2.8	Typically used wavelengths for SAR applications extracted from Flores-Anderson et al. [2019].	49
2.9	Water depth retrieval from flood extents with their respective accuracies and validation data.	59
3.1	Summary of symbols used in UQ/DA and their context in the thesis. . . .	64

Chapter 1

Introduction

Contents

1.1	General context	14
1.2	Industrial context	16
1.3	Scope of the thesis	17
1.4	Outline of the manuscript	18
1.5	Scientific productions	19

1.1 General context

Fluvial floods occur when water overflows the banks of a river or breaks through levees, resulting in water in adjacent floodplains. While these floods (especially morphogenic floods with a low return period of one to two years) are necessary for sustaining natural habitats, generating electricity, or supporting irrigation, they are also qualified as a negative natural hazard when excessive flow occurs. Major flood events (e.g., Western and Central Europe in July 2021, Pakistan in 2022, and Spain in October 2024) impact millions of people and result in economic losses of tens of billions of euros [Jongman et al., 2012]. The frequency and intensity of floods have increased due to climate change, and this phenomenon is expected to intensify [Zhongming et al., 2021]. Estimates suggest that the number of people exposed to flooding with an annual probability above 0.01 will rise from 800 million in 2010 to 1 billion in 2050 [Jongman et al., 2012]. Therefore, reducing the negative impacts and adapting to this phenomenon is essential.

Numerical models are valuable tools for estimating flood dynamics, including water depth, velocity, and flood extent, by solving fluid motion equations. These models are essential for flood forecasting, for generating representative distribution of flood scenarios with various inflows. Fluvial flood modeling typically relies on one-dimensional or two-dimensional numerical models solving the Shallow Water Equations. Due to advances in high-performance computing and the availability of detailed data, most flood modeling studies now rely on two-dimensional models to better capture floodplain flows. The water depth should be predicted with a precision of up to tens of centimeters at most, as small differences in predicted water depth can lead to significant changes in flood extent, particularly in areas with flat topography. Then, using high-precision and high-resolution topographic data is essential. Fewtrell et al. [2008] and Savage et al. [2016] advocate for flood models with a spatial resolution ranging from 2-10 m in urban environments and 10-100 m in rural floodplains.

Despite the significant improvement in computational resources and the accuracy of numerical models, epistemic uncertainties due in particular to the turbulent nature of waterflows limit the performance of hydraulic models. The primary sources of uncertainties in these models are: i) the model inputs (inflow at the upstream, initial conditions, topography, and bathymetric data), ii) the mesh discretization, iii) the accurate representation of physical processes, iv) the values of the roughness parameter, and v) the number and precision of observations of flood events to calibrate the model and validate its performance [Bates et al., 2014]. To enhance the forecasting capacity of these models, they are calibrated using observations of past events. The model’s input and parameters, including boundary conditions, turbulence models, roughness, and hydraulic conductivity, can be calibrated. While a range of parameters and inputs can be calibrated based on observations, the usual calibrated parameter is the spatial distribution of roughness (representing the resistance opposing the flow because of vegetation, buildings, skin roughness, etc.), impacting water depth and velocities (see Figure 1.1). Indeed, a common assumption is to use a fixed discretization and a predefined representation of physical processes. In instrumented areas, it is also typically assumed that model inputs, such as inflow hydrographs, topographic, and bathymetric data, are measured with low uncertainty.

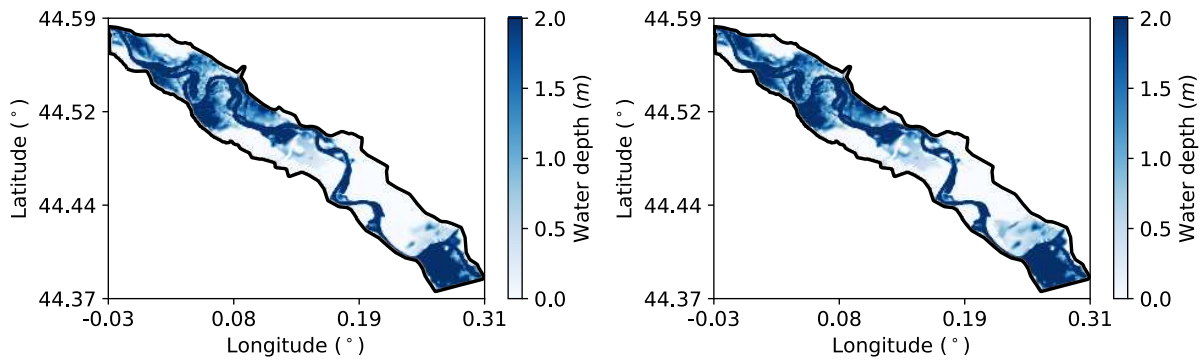


Figure 1.1: Water depth on a two-dimensional simulation with two spatial configurations of the roughness parameter.

Parameterizing and calibrating roughness in floodplains is challenging because it cannot be directly measured in the field. Consequently, the roughness and its spatial distribution are largely unknown beforehand. Although indirect measurements, expert judgments, and land use data constrain the range of the parameters based on tabulated values [Chow, 1959; Arcement and Schneider, 1989; Barnes, 1967], calibration is required to optimize the discrepancy between the model and observed data. Most factors (e.g., vegetation resistance, microtopography effects) contributing to the resistance to flow are unmeasurable. Therefore, the calibration of this parameter does not represent only the “true” physical value of the roughness but also compensates for the other errors in the input and validation data (e.g., observation errors, inlet discharge, topographic measurement errors, mismatch between the simplified model hydraulic processes and truth).

Riverbed roughness is typically calibrated using stage-gauging stations along the channel, which measure water depth regularly. However, these data are often sparse, especially in remote areas with ungauged catchments. Furthermore, the number of stage-gauging stations has declined in Europe since the 1980s [Crochemore et al., 2020], hindering our capacity to calibrate numerical models. While stage-gauging stations provide information to calibrate the roughness in the minor bed, few observations are available for the floodplains.

In the last five decades, the number of satellites orbiting the Earth has skyrocketed (around 2,200, with more than 150 dedicated to Earth Observation), providing a unique

opportunity to capture “photographs” of the Earth to monitor natural disasters such as floods (see Figure 1.2). Satellite imagery can be a crucial source of information for humanitarian relief and decision-makers during floods, providing near-real-time flooding maps and enhanced predictions for future floods [Voigt et al., 2016]. Emergency mechanisms initiated by major space agencies, such as the Copernicus Emergency Management Service of the European Commission [Copernicus, 2022], or the “Space and Major Disasters” [Bessis et al., 2004], provide satellite data and near real-time products for emergency relief during floods including the Global Flood Monitoring product, which is based on Sentinel-1 data integrated into the Global Flood Awareness (GloFAS) System [Salamon et al., 2021].

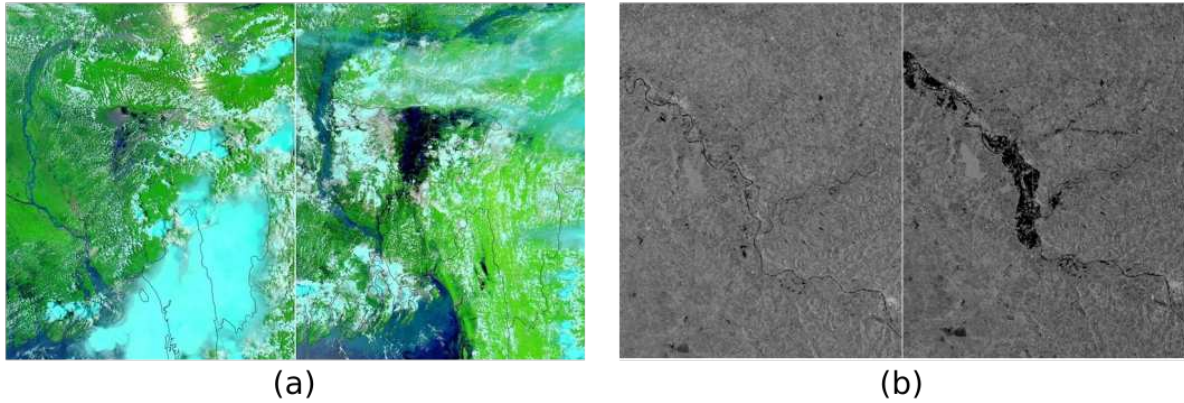


Figure 1.2: (a) Optical satellite image (MODIS mission) in Bangladesh before flood on June 5 (left) and during flood on July 25 (right) 2022, (b) Radar satellite image (Sentinel-1 mission) before flood on January 16 (left) and during on February 3 (right) 2021 over the Garonne River.

Satellite data can enhance both the calibration (e.g., the roughness of floodplains) and the validation of hydraulic models [Grimaldi et al., 2016; Wood et al., 2016]. This information includes raw satellite measurements [Cooper et al., 2019], point-wise water depth with altimeters [Domeneghetti et al., 2014, 2015], flood extent derived from optical or SAR satellites [Lai et al., 2014; Hostache et al., 2018; Dasgupta et al., 2021b], water depth extracted from flood extent [Hostache et al., 2009; Mason et al., 2009a], or interferometric measurements of water depth changes over time with SAR data [Alsdorf et al., 2001]. A key challenge is integrating satellite observations into the modeling process while accounting for their uncertainties. The aim is to find the best compromise between the prior values of the parameters on the floodplains and the information provided by these uncertain observations.

1.2 Industrial context

EDF is an international company that produces, distributes, and supplies electricity. EDF generates most of its electricity from nuclear power plants and renewable sources. Nuclear power plants are generally located by the sea or a river, providing access to a source of cold water for the cooling process via a pumping system. However, this also means that the plants are exposed to external floods (floods with origins outside the nuclear power plant installations). For over two decades, the LNHE, a department of EDF R&D, has been at the forefront of reinforcing nuclear safety against external floods.

An illustrative example of this challenge occurred during the Martin storm in December 1999, when part of the Blayais nuclear power plant was flooded. Following this event,

the flood risk for nuclear plants was requalified, and significant improvements were made, including raising the elevation of protective dikes, as shown in Figure 1.3.



Figure 1.3: Aerial view of the Blayais nuclear power plant with its four production units and the protective dikes. *Credit: Didier Marc.*

EDF must demonstrate the effectiveness of safeguards against the threat of external flooding at nuclear facilities. Since 2013, a new guide, prepared in collaboration between the French ASN, the operators of nuclear power plants, and research organizations, describes the doctrine of nuclear safety against external flooding [[Autorité de Sûreté Nucléaire, 2013](#)]. This guide defines eleven reference flooding scenarios and provides recommendations for evaluating these risks, particularly in the case of rare (usually with a frequency of occurrence of the order of 10^{-4}) but severe events. This guide proposes guidelines for characterizing these extreme phenomena while accounting for uncertainties related to the characterization of such hazards. For instance, the quantification of roughness in hydraulic models plays a crucial role in three of the eleven reference flooding scenarios: “Large watershed flooding” (CGB), “Sea level” (NMA), and “Failure of a water-retaining structure” (ROR).

In this context, the MOISE project aims to improve EDF’s approaches for characterizing extreme flooding phenomena. This includes developing the state of the art, identifying new methodologies, and applying them for safety demonstrations. This thesis is funded by the MOISE project.

1.3 Scope of the thesis

To ensure that the current ranges of floodplain roughness parameters used in flood engineering studies are accurate, calibration should be supported by the maximum amount of observed data. Satellite observations can be leveraged to extract flood extent maps or water depth fields (spatially distributed data) during a flood, and then used as observations for calibrating the model. This thesis aims to develop a methodology for characterizing the roughness coefficients of 2D hydrodynamic models to simulate extreme floods, using uncertain satellite observations. To achieve this, the manuscript focuses on answering the following research questions.

Research Questions

1. What information can be derived from satellite observations to calibrate 2D flood models, and how accurately can it be quantified?
2. How can the agreement between a satellite observation and a simulation be quantified?
3. Is it possible to calibrate floodplain roughness in 2D hydrodynamic models using uncertain satellite observations?

Various methods from the literature are tested to extract flood extent maps and water depth fields from satellite observations, while quantifying their uncertainties. The influence of satellite image preprocessing is evaluated, as well as the impact of method selection and hyperparameter settings on the accuracy of extracted flood maps and water depth estimates. The extracted flood extent maps from satellite observations should be compared with the outputs of the hydrodynamic model. Thus, we developed a step-by-step methodology to guide the selection of appropriate performance measures to compare simulated and observed flood extent maps. Then, DA techniques are employed to combine uncertain prior information on floodplain roughness with the uncertain satellite observations. The satellite observations (e.g., flood extent maps) are assimilated to quantify the posterior distribution of the roughness distribution. The objective is to evaluate the validity of current practices for parameterizing floodplain roughness. Specifically, the goal is to determine whether uncertain satellite observations and the extracted uncertain water depth fields or flood maps can be used to improve the quantification of uncertain roughness parameters in 2D flood models.

1.4 Outline of the manuscript

The manuscript is organized as follows:

- Chapter 2 introduces the governing equations of 2D flood modeling, especially the SWE, focusing on the roughness term that we aim to calibrate. This chapter also presents the theoretical background of satellite observations and their current applications in flood modeling.
- Chapter 3 describes the theoretical basis behind UQ and DA methods used in geosciences and in this manuscript.
- Chapter 4 provides an overview of flood mapping approaches for extracting flood extent and water depth from satellite observations. This study uses SAR imagery to evaluate various preprocessing (especially speckle noise reduction), flood mapping, and water depth estimation methods. The impact of the choice of method at the different steps and their hyperparameters are evaluated.
- Chapter 5 reports a step-by-step methodology to guide the selection of performance measures to compare numerical simulations of a flood event with a satellite image for model calibration or validation.
- Chapter 6 combines Chapters 4 and 5 insights to assimilate flood maps derived from satellite observations with various performance measures to calibrate the spatially varying roughness in a 2D flood model.

- Chapter 7 aims to assimilate water depth fields derived from satellite observations (instead of flood maps) to calibrate a spatially varying roughness field.
- Chapter 8 summarizes the main conclusions of the manuscript and proposes perspectives to enhance the integration of satellite observations in flood modeling studies.

1.5 Scientific productions

Several scientific contributions were made during this thesis, including published articles, conference papers, presentations, and a popularization article.

One article was published in Water Resources Research international peer-reviewed journal:

- **Travert J.-P.**, Boyaval S., Goeury C., Bacchi V., Zaoui F. (2025). Evaluation of performance measures for comparing flood models with satellite observations. Water Resources Research, 61, e2024WR038506 (<https://doi.org/10.1029/2024WR038506>).

Two articles are under revision in Natural Hazards and Earth System Sciences and Applied Ocean Research international journals:

- **Travert J.-P.**, Goeury C., Boyaval S., Bacchi V., Zaoui F. (2025). Evaluating the effects of preprocessing, method selection, and hyperparameter tuning on SAR-based flood mapping and water depth estimation.
- Thiébot J., Goeury C., **Travert J.-P.**, Sebastian A., Salomon J. (2025). The contribution of sensitivity analysis and data assimilation to tidal-stream resource assessment: the example of the Alderney Race, Under Review

Two articles were submitted:

- **Travert J.-P.**, Boyaval S., Goeury C., Bacchi V., Zaoui F. (2025). Data assimilation of flood maps in a 2D flood model for different performance measures
- Sebastian A., Thiébot J., Goeury C., **Travert J.-P.**, Poizot E., Gregoire G. (2025). Automatic Calibration and Simultaneous Estimation of Bottom Friction and Open Boundary Conditions in a Tidal Model

Two conference papers were submitted and presented at the Telemac User Conference international meeting in 2023 and 2024:

- **Travert J.-P.**, Goeury C., Bacchi V., Zaoui F., Boyaval S. (2023) Comparing data assimilation algorithms for parameter estimation on TELEMAC-2D, Proceedings of the XXIXth TELEMAC Users Conference 12-13 October 2023
- **Travert J.-P.**, Goeury C., Bacchi V., Zaoui F., Boyaval S. (2024) Influence of 3D-VAR Dissimilarity Measure on Flood Model Calibration with Satellite Data, Proceedings of the XXXth TELEMAC Users Conference 08-10 October 2024

Two posters were presented during international meetings:

- **Travert J.-P.**, Goeury C., Bacchi V., Zaoui F. and Boyaval S. (2023) Flood twin experiment for estimating the potential of satellite observations in shallow-water simulations, EGU General Assembly Conference Abstracts, pp. EGU-17379.

- **Travert J.-P.**, Goeury C., Bacchi V., Zaoui F. and Boyaval S. (2023) Flood twin experiment for evaluating the potential of satellite data in shallow-water simulations, 40th IAHR World Congress 'Rivers - Connecting Mountains and Coasts' 21-25th August 2023

Two presentations were delivered at the SimHydro and IAHR Europe conferences:

- **Travert J.-P.**, Goeury C., Bacchi V., Zaoui F. and Boyaval S. (2023) Mapping of roughness zones in 2-D hydraulic models with satellite observations, SimHydro 2023
- **Travert J.-P.**, Goeury C., Bacchi V., Zaoui F. and Boyaval S. (2023) Parameterization of Roughness Distribution on Floodplains in 2D Shallow-Water Simulations, 8th IAHR Europe Congress, Lisbon 4-7 June 2024
This presentation received a special mention for the Gerhard Jirka Award for Young Researchers 2024, recognizing the best presentation given by a Young Professional during the biennial IAHR Europe Division Congress (<https://www.iahr.org/index/detail/1400>).

A popularization article was co-authored with Guillaume Coulaud (PhD student at EDF R&D), focusing on the protection of nuclear power plants from flooding events:

- “Numerical simulation of flood risks in nuclear power plants”
<https://ingenius.ecoledesponts.fr/en/articles/numerical-simulation-of-flood-risks-in-nuclear-power-plants/>

Chapter 2

Flood Modeling and Satellite Observations

Résumé :

Ce chapitre présente des approches communes de modélisation numérique pour la modélisation des inondations. Les équations utilisées pour modéliser les écoulements à surface libre, y compris les équations de Navier-Stokes et les équations de Saint-Venant sont présentées. Les paramètres du modèle doivent être calibrés pour correspondre aux observations. Une attention particulière est accordée au terme de rugosité dans les équations de Saint-Venant, qui est généralement soumis à une calibration.

Avec la disponibilité croissante des observations par satellite, les données spatialement distribuées peuvent maintenant aider à la surveillance des inondations. Ainsi, le chapitre présente ensuite les principes fondamentaux de l'observation satellitaire en mettant l'accent sur les satellites radar à synthèse d'ouverture qui sont couramment utilisés pour extraire des informations pour calibrer les modèles d'inondation. Les principales caractéristiques et limites de l'imagerie radar à synthèse d'ouverture sont présentées, en mettant l'accent sur l'interaction entre le signal radar et les surfaces inondées, ainsi que sur les types d'informations qui peuvent être extraites de ces observations et les méthodes utilisées à cette fin.

Abstract:

This chapter presents common numerical modeling approaches for flood modeling. The governing equations used to model free-surface flows, including the Navier-Stokes equations and SWE, are presented. Model parameters must be calibrated to fit the observations. Particular attention is given to the roughness term in the SWE, which is typically subject to calibration.

With the increasing availability of satellite observations, spatially distributed data can now support flood monitoring and management. Thus, the chapter then introduces the fundamental principles of Earth Observation, with a specific focus on SAR satellites, which are commonly used to generate information for calibrating flood models. The main characteristics and limitations of SAR imagery are discussed, with a focus on the interaction between the radar signal and water surfaces, as well as the types of information that can be extracted from these observations and the methods used to do so.

Contents

2.1	Modeling free-surface flows	23
2.1.1	Hydrodynamic equations	23
2.1.2	Modeling system	27
2.1.3	Bed shear stress modeling	28
2.2	Monitoring floods with satellite observations	37
2.2.1	Electromagnetic spectrum	38
2.2.2	Satellite sensors	39
2.2.3	A focus on Synthetic Aperture Radar	42
2.2.4	Interaction of SAR with the Earth's surface	47
2.2.5	Detection of water surfaces with SAR observations	52
2.3	Summary	59

2.1 Modeling free-surface flows

Since the 19th century, flood monitoring has relied heavily on empirical observations and simplified hydraulic formulas to estimate the flow dynamics. The motions of fluids are described with hydrodynamic equations derived from the laws of conservation of mass and momentum. The Navier-Stokes equations were formulated to describe the motion of fluids in three dimensions. However, because of the complexity of solving the Navier-Stokes equations, especially with a moving domain with a free surface, simplified 2D models such as the SWE are often preferred for industrial applications and large-scale problems. A reminder of the Navier-Stokes equations and SWE is proposed in 2.1.1. These equations are then solved with a modeling system described in Section 2.1.2. In the SWE, the bottom shear stress term is usually calibrated to fit the numerical models to observations. More details on this term can be found in Section 2.1.3, including an overview of its quantification with and without observations.

2.1.1 Hydrodynamic equations

This section describes the mathematical models governing the motion of water, particularly free surface flows [Chow, 1959; Thual, 2010]. First, the Navier-Stokes equations are presented. Then, we describe the hypotheses and the resulting 2D equations from the Navier-Stokes equations. For a more in-depth look at this topic and details on the assumptions and derivation of the equations, readers can refer to books and courses such as Gerbeau and Perthame [2000], or Thual [2010].

2.1.1.1 Navier-Stokes equations

The Navier-Stokes equations are fundamental in fluid dynamics, describing the motion of fluids (liquids, gases, and plasmas) in 3D space. These equations were formulated independently by Louis Marie Henri Navier and George Gabriel Stokes at the beginning of the 19th century.

Transport theorem of Reynolds Let us consider a small domain of fluid Ω with surface Σ , on which the volume integral of a physical quantity, such as the mass or momentum, is computed. We denote by $I(t)$ the integral of the continuous function $A(\mathbf{X}, t)$ representing a given quantity on the control volume at position $\mathbf{X} = [x, y, z]$ and time t :

$$I(t) = \int_{\Omega(t)} A(\mathbf{X}, t) d\Omega. \quad (2.1)$$

The temporal variation of I is given by the transport theorem as follows:

$$\frac{dI(t)}{dt} = \frac{d}{dt} \left(\int_{\Omega(t)} A(\mathbf{X}, t) d\Omega \right) = \int_{\Omega(t)} \frac{\partial A(\mathbf{X}, t)}{\partial t} d\Omega + \int_{\Sigma(t)} A(\mathbf{X}, t) (\mathbf{U}_\Sigma \cdot \mathbf{n}) dS \quad (2.2)$$

where $\mathbf{U}_\Sigma$ is the velocity on the surface, and $\mathbf{n}$ is the normal to this surface, oriented outwards.

Continuity equation Let us apply the transport theorem from Equation (2.2) to the density ρ , with velocity $\mathbf{U} = [U, V, W]$ over a fluid domain Ω with surface Σ :

$$\frac{d}{dt} \int_{\Omega} \rho(\mathbf{x}, t) d\Omega = \int_{\Omega} \frac{\partial \rho}{\partial t} d\Omega + \int_{\Sigma} \rho \mathbf{U} \cdot \mathbf{n} dS. \quad (2.3)$$

The flux across the domain boundaries $\int_{\Sigma} \rho \mathbf{U} \cdot \mathbf{n} dS$ can be rewritten as $\int_{\Omega} \nabla \cdot (\rho \mathbf{U}) d\Omega$ using Stokes' theorem. When considering mass conservation on Ω , then $\frac{d}{dt} \int_{\Omega} \rho(\mathbf{x}, t) d\Omega = 0$, and by considering a tiny volume of fluid, we can write a local formulation such as:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{U}) = 0. \quad (2.4)$$

Momentum conservation By applying the transport theorem from Equation (2.2) to momentum $\rho \mathbf{U}$, we obtain:

$$\frac{d}{dt} \int_{\Omega} \rho \mathbf{U} d\Omega = \int_{\Omega} \left(\frac{\partial \rho \mathbf{U}}{\partial t} + \nabla \cdot (\rho \mathbf{U}) \right) d\Omega. \quad (2.5)$$

The fundamental principle of dynamics (Newton's second law) states that the changes in momentum are equal to the external forces applied to a control volume, then:

$$\frac{d}{dt} \int_{\Omega} \rho \mathbf{U} d\Omega = \int_{\Omega} (\rho \mathbf{g} + \mathbf{F}) d\Omega + \int_{\Sigma} \underline{\underline{\sigma}} \mathbf{n} dS = \int_{\Omega} (\rho \mathbf{g} + \mathbf{F} + \nabla \cdot \underline{\underline{\sigma}}) d\Omega, \quad (2.6)$$

where $\underline{\underline{\sigma}}$ is the viscous stress tensor, $\rho \mathbf{g}$ is the volumetric gravity forces, and $\mathbf{F}$ is the volumetric external forces. Finally using Equation (2.4), local formulation writes as:

$$\rho \left(\frac{\partial (\mathbf{U})}{\partial t} + (\mathbf{U} \cdot \nabla) \mathbf{U} \right) = \nabla \cdot \underline{\underline{\sigma}} + \rho \mathbf{g} + \rho \mathbf{F}. \quad (2.7)$$

Most fluids in the environment, including water, can be considered Newtonian, as the viscous stress is linearly proportional to the deformation rate. For Newtonian fluids, Stokes' law expresses the viscous strain tensor as:

$$\underline{\underline{\sigma}} = -p \mathbf{I} + \lambda \nabla \cdot \mathbf{U} \mathbf{I} + 2\mu \mathbf{s}, \quad (2.8)$$

where p is the pressure, $\mathbf{I}$ is the identity tensor, μ is the dynamic molecular viscosity, $\lambda = -\frac{2}{3}\mu$ with Stokes hypothesis [Stokes, 1845], and $\mathbf{s}$ is the strain-rate tensor. The strain-rate tensor writes as:

$$\mathbf{s} = \frac{1}{2} (\nabla \mathbf{U} + \nabla \mathbf{U}^T). \quad (2.9)$$

From this set of equations, the Navier-Stokes equations for a Newtonian and incompressible fluid ($\nabla \cdot \mathbf{U} = 0$) are derived:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{U}) = 0 \\ \frac{\partial (\mathbf{U})}{\partial t} + (\mathbf{U} \cdot \nabla) \mathbf{U} = -\frac{1}{\rho} \nabla p + \nu \Delta \mathbf{U} + \mathbf{g} + \mathbf{F} \end{cases}, \quad (2.10)$$

where $\nu = \frac{\mu}{\rho}$ is the kinematic viscosity. In turbulent flows, fluid particles undergo chaotic small-scale motions that increase momentum transfer. That transfer can be modeled by an effective viscosity (e.g., as described by Boussinesq). To account for this additional dissipation, caused by vortices and instationarities in the flow, a turbulent viscosity term denoted as ν_t is often added to the kinematic viscosity. To solve the equations involving turbulent viscosity, which is a function of additional turbulent-related variables, a closure approach is required. Closure refers to the process of introducing additional equations or models that express the turbulent-related quantities in terms of the other variables, thereby making the system of equations solvable.

Navier-Stokes equations should be completed to account for the problem's geometry and to close the system. For river and flood modeling, we consider that:

- The bottom and lateral boundaries (dikes, walls, etc.) are impermeable.
- The fluid particles at the free surface remain part of the free surface.
- Inlet and outlet boundaries of the domain are prescribed with additional information such as velocities and/or water depth depending on the flow regime. For subcritical flows, the discharge or the water depth should be imposed at the inlet and outlet. For supercritical flows, the discharge and the water depth are specified at the inlet, but no outlet conditions are required.
- To close the Navier-Stokes equations, one needs to define the surface and bottom shear stresses.

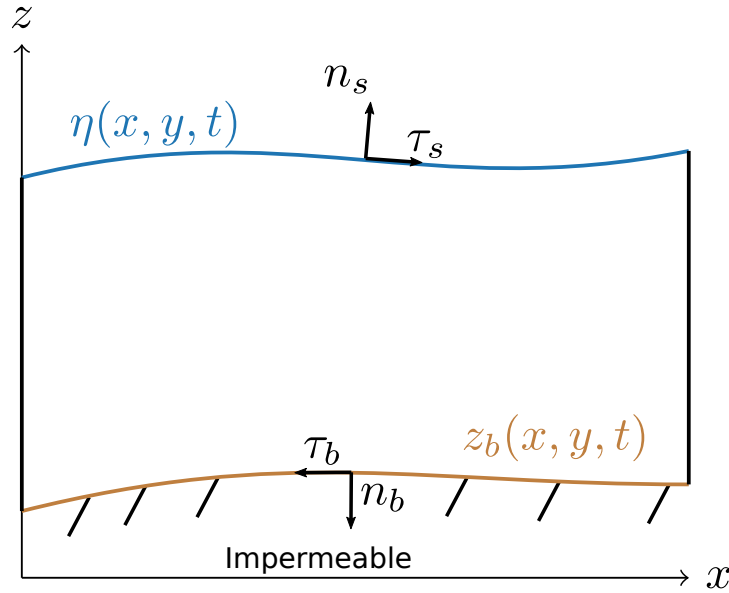


Figure 2.1: Schematic of the free surface, bottom, and shear stresses.

Free surface condition: The particles at the free surface should remain at the free surface over time. Let us consider the quantity $\Phi(x, y, t) = z - \eta(x, y, t)$ with z as the elevation in the vertical direction and η the free surface elevation (see Figure 2.1). On the free surface, $\frac{d\Phi}{dt}$ should equal zero so that the particles stay on the free surface. Therefore, applying the transport theorem leads to the relation known as the kinematic condition:

$$\frac{\partial \Phi(x, y, t)}{\partial t} + \mathbf{U} \cdot \nabla \Phi(x, y, t) = 0. \quad (2.11)$$

Shear stresses: Equation (2.10) should be completed with shear stresses at the bottom (e.g., caused by vegetation) and at the surface (e.g., caused by the wind), as illustrated in Figure 2.1. The bottom shear stresses are defined as $\boldsymbol{\tau}_b = \underline{\underline{\sigma}} \cdot \mathbf{n}_b$ and $\boldsymbol{\tau}_s = \underline{\underline{\sigma}} \cdot \mathbf{n}_s$ where $\mathbf{n}_b$ and $\mathbf{n}_s$ are the outward unit normal at bottom and surface boundaries, respectively. Free-surface constraints $\boldsymbol{\tau}_s$ are usually neglected for river floods (for ocean modeling, the free-surface constraint can affect wave propagation). For the bottom shear stress, a classical closure model is:

$$\boldsymbol{\tau}_b = \frac{1}{2} \rho C_f \|\mathbf{U}\| \mathbf{U}, \quad (2.12)$$

with C_f a dimensionless friction coefficient function of the bottom roughness modeled with empirical or semi-empirical laws [Morvan et al., 2008]. These laws will be discussed in detail in Section 2.1.3.

Note: The bottom shear stress τ_b depends on parameters (according to the closure model) tabulated in the literature or calibrated based on observations. This parameter is usually calibrated in flood studies and is the main focus of the thesis. The bottom elevation z_b and the inlet discharge are considered to be perfectly known.

2.1.1.2 Shallow Water Equations

The numerical resolution of the Navier-Stokes equations is costly for large domains (kilometers) requiring fine resolution, especially when in a moving domain whose time evaluation has to be computed with Equation (2.11) and that matrices should be reassembled at each time step. Furthermore, the uncertainties due to “effective viscosity” make the 3D calculations even more uncertain than the 2D ones. Thus, flood modelers typically prefer to use 2D numerical models to simulate horizontal flows. The SWE, also known as the Saint-Venant equations, were introduced in 1871 and named after Adhémar Jean Claude Barré de Saint-Venant [de Saint-Venant, 1871]. The derivation of these equations from the Navier-Stokes equations for incompressible flows is detailed in Gerbeau and Perthame [2000].

Hypotheses

- The fluid is incompressible, homogeneous, and Newtonian.
- The vertical scale (the water depth (h)) is small compared to the characteristic horizontal scale (L), that is to say, $\frac{h}{L} \ll 1$. The velocity is considered constant in shallow waters within the water column. The depth-averaged velocities are defined as:

$$\begin{cases} u(x, y, t) = \frac{1}{h(x, y, t)} \int_{z_b(x, y, t)}^{\eta(x, y, t)} U(\mathbf{X}, t) dz \\ v(x, y, t) = \frac{1}{h(x, y, t)} \int_{z_b(x, y, t)}^{\eta(x, y, t)} V(\mathbf{X}, t) dz \end{cases}, \quad (2.13)$$

where $h = \eta(x, y, t) - z_b(x, y, t)$, η is the surface elevation and z_b is the bottom elevation.

- Negligible vertical velocities (w) relatively to horizontal velocities u and v . This assumption is also necessary for the hydrostatic pressure hypothesis. In river channels where the width is at least ten times larger than the water depth, this assumption remains valid. The approximation is accurate even for oceans with significant depths because their horizontal extent (thousands of kilometers) far exceeds their vertical extent (a few kilometers).
- The pressure at any point in the fluid depends only on the depth below the surface, such that $p = \rho gh$ (hydrostatic pressure).
- The bottom slope is shallow, usually the slope is below 10^{-3} m/m or 10^{-4} m/m.

Continuity equation From the integration of Equation (2.4) over the vertical and using the Leibniz rule, we obtain:

$$\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) = 0, \quad (2.14)$$

where $\mathbf{u} = (u(x, y, t), v(x, y, t))$ is the velocity vector in 2D without the vertical dimension. Comparing this equation with the mass conservation of Navier-Stokes shows that water depth plays the role of water density.

Momentum conservation From the vertical component of the Navier-Stokes momentum equation, and considering that external forces are constant along the vertical, we derive the hydrostatic pressure $\frac{\partial p}{\partial z} = -\rho g$. Finally, by applying depth-averaging on the vertical for the remaining equations with hydrostatic pressure and kinematic condition, the SWE derive as:

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} = 0 \\ \frac{\partial(hu)}{\partial t} + \frac{\partial(hu^2)}{\partial x} + \frac{\partial(huv)}{\partial y} = -gh \frac{\partial \eta}{\partial x} + h F_x + \nabla \cdot (h \nu_e \nabla u) - \frac{1}{\rho} \tau_{bx} \quad , \\ \frac{\partial(hv)}{\partial t} + \frac{\partial(hv^2)}{\partial y} + \frac{\partial(huv)}{\partial x} = -gh \frac{\partial \eta}{\partial y} + h F_y + \nabla \cdot (h \nu_e \nabla v) - \frac{1}{\rho} \tau_{by} \end{cases} \quad (2.15)$$

where ν_e is the effective viscosity, which is the sum of diffusion expressed by kinematic and turbulent viscosity, and the dispersion in ν_d . The dispersive viscosity ν_d arises from approximating the integration of the non-linear term.

Note: In the manuscript only the SWE are used, as 2D hydrodynamic models are generally sufficient for modeling flood flows. However, the Navier-Stokes equations were introduced to explicitly identify the physical simplifications involved in the derivation of SWE. During model calibration, these simplifications may be partially compensated for by adjusting model parameters.

2.1.2 Modeling system

The equations listed in Section 2.1.1 are implemented within the open-source open-TELEMAC system (www.opentelemac.org), developed primarily by the LNHE department of Electricité de France's Research and Development Division since 1987 [Hervouet, 2007]. The code solves complex free-surface flow problems for environmental purposes. As illustrated in Figure 2.2, the openTELEMAC system consists of different interconnected modules for hydraulics (MASCARET, TELEMAC-2D, and TELEMAC-3D), waves (ARTEMIS and TOMAWAC), water quality (WAQTEL), frazil (KHIONE), and sediments (COURLIS and GAIA) modeling. The openTELEMAC system is used for various applications, such as coastal and marine (overtopping on dikes, tsunamis, and oil spills), fluvial (floods, dike breaches, tracer transport), and other more specific cases. The code solves regional to local-scale problems, and has parallelization capabilities for computationally intensive simulations. For modeling floods, we resort to the T2D model, which solves SWE as defined in Equation (2.15). This enables the calculation of water depth and velocity field.

Note: All the simulations in the manuscript are carried out with Finite Elements, based on an unstructured triangular mesh. The N edge-by-edge scheme and the Positive Streamwise Invariant Distributive scheme are used to compute the advection of velocity and water depth, respectively. The wetting and drying of grid elements, which are essential in floodplain modeling, are managed by correcting the free surface gradient. More details on the Finite Elements numerical schemes are found in [Hervouet et al. \[2017\]](#). These considerations are the standard choices for flood modeling studies at EDF, typically applied to large watershed flooding or dam break floods [[Bacchi and Pavan, 2025](#)].

One of the strengths of the openTELEMAC system is its ability to interact with the code during simulations through APIs implementations for each module. The APIs provide flexibility by allowing users to control the simulations while they are running [[Goeury et al., 2022](#)]. Users can stop the simulation, fetch specific variable values, and potentially modify them before continuing the computation by offering direct access to the physical memory of variables. The APIs are not only available within a Fortran framework, but the openTELEMAC system also contains a package called TelApy, which is a Python wrapper for the APIs providing a user-friendly language and creating a more efficient calculation chain.

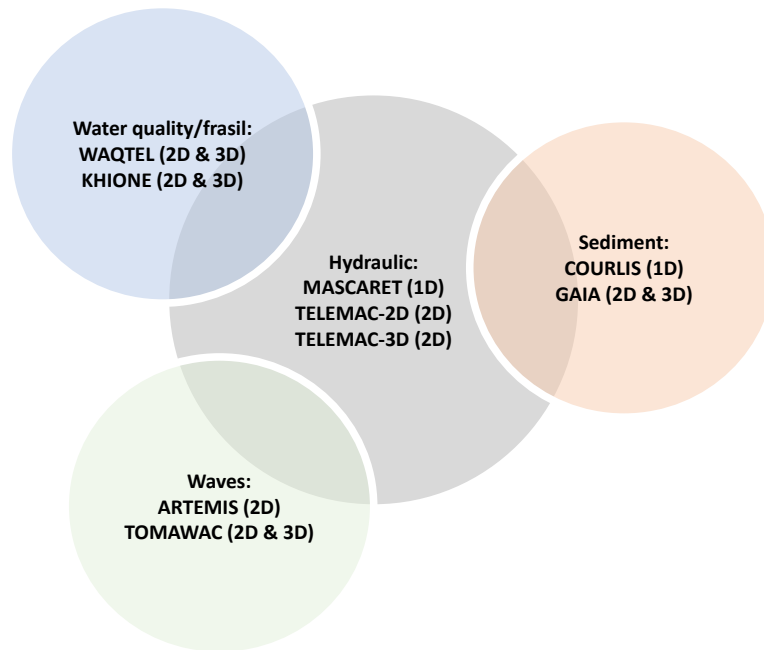


Figure 2.2: openTELEMAC architecture adapted from [Goeury et al. \[2022\]](#).

2.1.3 Bed shear stress modeling

As introduced in Section 2.1.1, a crucial component of flood modeling is the bed shear stress τ_b , commonly referred to as roughness or the friction term. This section presents details about this term, its physical origins, and its significance in hydraulic simulations.

2.1.3.1 Physical explanation of roughness

Roughness represents the resistance to water flow caused by the boundaries of a channel. A channel with high roughness introduces substantial friction, leading to considerable

energy losses that decelerate the water flow while simultaneously elevating the water level. Conversely, for a smooth channel surface, energy losses are minimal, allowing water to flow more rapidly. The bottom shear stress is usually not uniform or constant over time. It varies according to the hydraulic conditions and the evolving state of the riverbed. A range of physical factors influencing its magnitude is provided in [Chow \[1959\]](#) and described hereafter:

- **Surface roughness:** Surface roughness refers to the irregularities on the bottom of the flow. It corresponds to the size and shape of the material at the bottom, producing turbulence effects and changing the wetted perimeter.
- **Land use and vegetation:** Land use and vegetation act as surface roughness that opposes the flow. Buildings or forests may retard the flow, which depends heavily on the density, type of land use and vegetation, height, and their distributions.
- **Channel irregularities:** Irregularities such as sand bars, ridges, or holes perturb the flow.
- **Obstructions:** The presence of obstructions such as bridges, log jams.
- **State and discharge:** Water depth and velocity modifications, hence, modifying bedforms, and activating or suppressing sediment transport.
- **Seasonal change:** Changes in land use conditions with plant development, trees on the channel or the banks.
- **Channel geometry:** The curvature and size of the channel.
- **Suspended material or bedload:** Head losses due to bedload and/or suspended sediments in the water column.

Surface roughness and land use are the most influential contributors to friction. However, other processes, such as vegetation development, may contribute to friction and increase or reduce roughness.

2.1.3.2 Closure models for τ_b

The bed shear stress should be written as a function of flow velocities and water depth to close the SWE. In T2D, the friction term is considered in the momentum equation from the SWE and is treated implicitly. The two components of bottom friction are expressed as:

$$\begin{cases} \tau_{bx} = \frac{u}{2} \rho C_f \sqrt{u^2 + v^2} \\ \tau_{by} = \frac{v}{2} \rho C_f \sqrt{u^2 + v^2} \end{cases}, \quad (2.16)$$

where C_f is dimensionless, and its value depends on the chosen friction model. In the literature, many friction models are available to compute C_f based on empirical or semi-empirical approaches [[Morvan et al., 2008](#)]. In most flood modeling studies, the C_f coefficient is computed based on the Manning-Strickler formula. This formula is generally used because it is easy to implement with a single parameter to tune, and its value is well-documented in the literature for various conditions, as it has been calibrated for a wide variety of rivers [[Barnes, 1967](#); [Chow, 1959](#)].

Manning-Strickler formula In 1889, Robert Manning and Albert Strickler developed the widely used Manning-Strickler formula [Manning et al., 1890]. This formula is derived empirically from seven previous formulas based on Bazin’s experimental data and 170 additional observations such that:

$$C_f = \frac{g}{K_s^2 h^{1/3}}$$

$$\tau_{bx} = \tau_{by} = \frac{\rho g}{K_s^2 h^{7/3}} Q^2, \quad (2.17)$$

with $Q = h||\mathbf{u}||$ the discharge, K_s representing the Strickler coefficient with $n = \frac{1}{K_s}$ the Manning coefficient.

Note: In this thesis, we consider the Manning-Strickler formula and aim to calibrate the K_s parameter in the floodplains, such that $K_s(x, y)$ varies in space but is independent of time. This is the usual methodology used for flood modeling studies at EDF [Bacchi and Pavan, 2025]. In the literature, depending on the choice of model used to compute C_f , it has been shown that modeled water depths and velocities can vary slightly [Dalledonne et al., 2019]. The choice of the closure model is a first assumption in the calibration study, which could be questioned in the future. However, for flood flows, the calibration of the parameters is more influential than the choice of the closure model. The concept of roughness is difficult to describe, as it results from several physical factors. Therefore, the parameter (here K_s) is calibrated to model all physical processes near the bed [Morvan et al., 2008] and compensate for simplifications (e.g., 3D flow structures, unmodeled turbulence) and errors (e.g, bathymetry) in the model. We should also avoid non-physical outputs, so the parameter is usually bounded based on expert opinions and experiment setups as described in Section 2.1.3.3.

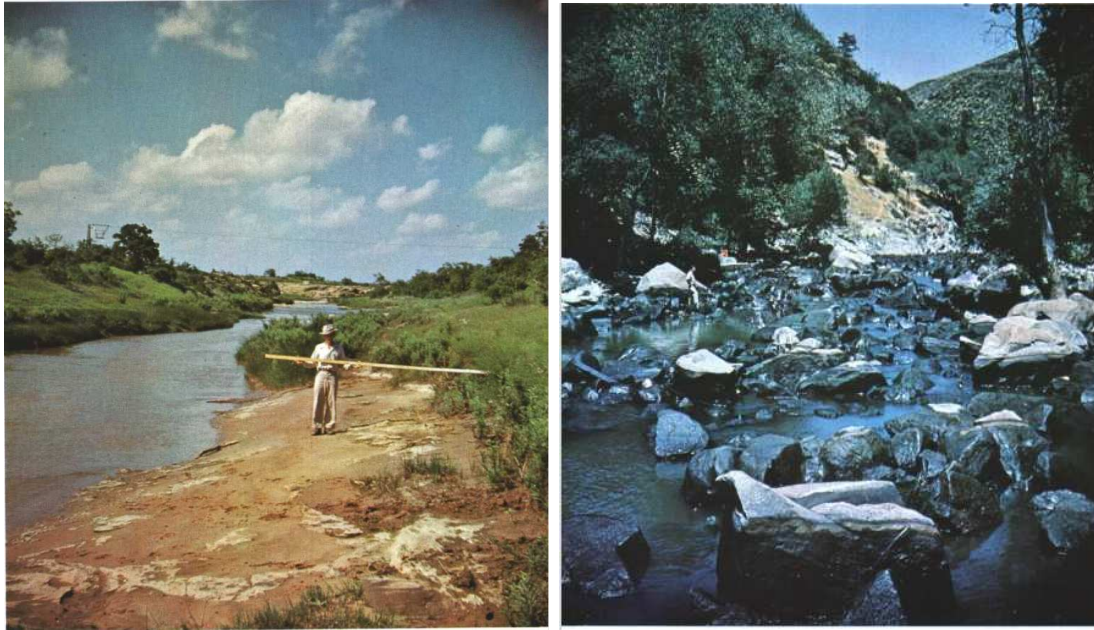
2.1.3.3 Roughness parameterization: Prior to calibration

To simulate flood events, the parameters in the closure model must be specified. Ideally, these parameters are estimated through calibration and validation based on historical flood events by adjusting the parameters to minimize the discrepancy between model simulations and observations. However, observations of water depth and velocity during floods are often limited in both time and space, and may be subject to uncertainty. Furthermore, the parameters must be constrained between bounds to avoid non-physical calibration. Consequently, parameter values are informed by prior knowledge such as expert judgment, laboratory experiments, or tabulated values.

Quantification based on expert opinions/literature, and in-situ measurements

As described in Section 2.1.3.1, the resistance of the flow is connected to the material of the bed. The estimation of this value for each material has been widely studied in several empirical studies [Chow, 1959; Arcement and Schneider, 1989; French, 1985], with most of the literature based on 1D approaches. These historical approaches rely on calibrated photographs and tables of roughness values for different material configurations [Barnes, 1967; Chow, 1959]. The photograph method is based on a series of photographs of stream channels that represent various roughness values calculated using the Manning-Strickler formula. Water levels and discharges were measured for each photograph, and topographical data of the riverbed and banks were collected to compute the Manning value. For example, Figure 2.3 reports two calculated Manning values for two distinct types

of natural streams. Trained engineers can estimate roughness channels within $\pm 15\%$ by comparing their study area with images of channels of known roughness.



(a) Champlin Creek near Colorado City (US) with $n = 0.027 \text{ s} \cdot \text{m}^{-1/3}$.

(b) Cache Creek near Lower Lake (US) with $n = 0.079 \text{ s} \cdot \text{m}^{-1/3}$.

Figure 2.3: Examples of photographs of stream channels with associated estimated Manning (n) from the Manning-Strickler formula. Credit: U.S. Geological Survey.

Chow [1959] presented a comprehensive table of Manning values, drawing data from various sources [Horton, 1916; Ramser et al., 1929]. The proposed values for channels and floodplains, respectively, are reported in Tables 2.1 and 2.2.

Table 2.1: Proposed Manning values for river channels.

Type of channel	Minimum	Mean	Maximum
1. Main Channels for minor streams (<30 m)			
a. Clean, straight, full stage, no rifts or deep pools	0.025	0.030	0.033
b. Same as above, but more stones and weeds	0.030	0.035	0.040
c. Clean, winding, some pools and shoals	0.033	0.040	0.045
d. Same as above, but some weeds and stones	0.035	0.045	0.050
e. Same as above, lower stages, more ineffective slopes and sections	0.040	0.048	0.055
f. Same as “d” with more stones	0.045	0.050	0.060
g. Sluggish reaches, weedy, deep pools	0.050	0.070	0.080
h. Very weedy reaches, deep pools, or floodways with heavy stands of timber and underbrush	0.075	0.100	0.150
i. Mountain streams with gravels, cobbles, and few boulders	0.030	0.040	0.050
j. Mountain streams with cobbles and large boulders	0.040	0.050	0.070
2. Main Channels for major streams (>30 m)			
a. Regular section with no boulders or brush	0.025	...	0.060
b. Irregular and rough intersection	0.035	...	0.100

Usually, Strickler coefficients vary between 25 and $45 \text{ m}^{1/3} \cdot \text{s}^{-1}$ for natural streams and between 50 and $80 \text{ m}^{1/3} \cdot \text{s}^{-1}$ for artificial streams depending on the covering material. For the floodplains, as the water flows face obstacles, the Strickler coefficients are usually lower

(leading to lower flow velocities and higher water depth). While these tables are handy, their effectiveness relies on the user's experience and subjectivity in correctly identifying or comparing land use types with photographs.

Table 2.2: Proposed Manning values for floodplains.

Type of floodplain	Minimum	Mean	Maximum
1. Pasture, no brush			
a. Short grass	0.025	0.030	0.035
b. High grass	0.040	0.050	0.070
2. Cultivated areas			
a. No crop	0.020	0.030	0.040
b. Mature row crops	0.025	0.035	0.045
c. Mature field crops	0.030	0.040	0.050
3. Brush	0.050	0.070	0.080
a. Scattered brush, heavy weeds	0.035	0.050	0.070
b. Light brush and trees, in winter	0.035	0.050	0.060
c. Light brush and trees, in summer	0.040	0.060	0.080
d. Medium to dense brush, in winter	0.045	0.070	0.110
e. Medium to dense brush in summer	0.070	0.100	0.160
4. Trees			
a. Dense willows, summer, straight	0.110	0.150	0.200
b. Cleared land, but with heavy growth and sprouts	0.030	0.040	0.050
c. Same as above, but with heavy growth of sprouts	0.050	0.060	0.080
d. Heavy stand of timber, a few down trees, little undergrowth, flood stage below branches	0.080	0.100	0.120
e. Same as above, but with flood stage reaching branches	0.100	0.120	0.160

Cowan [1956] suggested a physics-informed adjustment of the coefficients, considering various characteristics that could influence the roughness values. The Manning coefficient is then expressed as:

$$n = (n_0 + n_1 + n_2 + n_3 + n_4)m_5, \quad (2.18)$$

where n_0 is the value for straight, uniform, smooth channels in the natural materials involved, n_1 is an adjustment parameter for surface irregularities, n_2 is for adjusting the coefficient based on variations in the shape and size of the channel cross-section, n_3 is accounting for obstructions, n_4 is for vegetation and flow conditions and m_5 is a correcting factor for meandering channels. Tables defining the values of this set of five parameters are provided in Chow [1959].

Other empirically based formulas have been proposed to define the roughness value based on in-situ measurements [Marriott and Jayaratne, 2010]. The Strickler formula states that $n = \frac{d_{50}^{1/6}}{21.6}$ [Strickler, 1923] (applicable for slopes between 0.004 m/m and 0.025 m/m and hydraulic radius between 0.1 m and 10 m) and the Meyer-Peter formula states that $n = \frac{d_{90}^{1/6}}{8.2\sqrt{g}}$ [Meyer-Peter and Müller, 1948]. d_{50} refers to the median grain size diameter at which 50% of the sediment particles are finer, and d_{90} refers to the particle size diameter below which 90% of the sediments are finer. Other formulas are available in the literature, and choosing a formula should rely on various factors, such as the granulometry of surface roughness or ripples. For interested readers, many formulas have been developed for addressing the special case of submerged and emerged vegetation, with a review available in Sellin et al. [2003].

Remark: The roughness values are often considered constant over time during a simulated flood event, but for long-term simulations, it may be worthwhile to consider their temporal evolution, as it can change throughout the year with the development of vegetation due to seasonal or geomorphological changes [De Doncker et al., 2009].

While these empirical formulas provide a prior estimate of the roughness value, their uncertainty is poorly quantified. The uncertainty in determining prior roughness parameterization with in-situ measurements and visual interpretation was described in Wohl [1998], by estimating the Manning value of 5 rivers using four methods listed in Table 2.3 (visual estimation, Jarret’s formula [Jarrett, 1984], Cowan’s method, and Limerinos’ formula [Limerinos, 1970]), with large variability between the methods.

Table 2.3: Estimates of Manning’s n for five rivers with four estimation methods from Wohl [1998].

Method/River	Proudre	Arun	Chayang	Fitzroy	Paran
Jarret	0.024	0.051	0.096	0.014	0.034
Limerinos	0.037	0.069	0.104	0.019	0.047
Visual	0.045	0.060	0.090	0.030	0.045
Cowan	0.050	0.060	0.070	0.030	0.050

Medeiros et al. [2012] compared the computed values of the roughness parameters at 24 sites in Florida (US) based on terrain measurements and those extracted from land cover data with a representative set of all the land cover. The research demonstrated that the surface roughness derived from land use (with tables) and the one computed from field measurements had a RMSE of 0.083, leading to the characterization of sand as bare rock. There may have been errors in the on-site measurements and the land cover classification from 1992, which was not as precise as today. Nonetheless, uncertainties remain significant when using the tables directly, affecting values for each floodplain type or channel configuration.

Remark: In the absence of an alternative, these uncertain tabulated values are used to define bounds for the Strickler parameter depending on the type of floodplain. Hence it advocates for using as much as available data for calibrating these parameters and characterize the range of variations of the model outputs with UQ techniques (see Chapter 3 for theoretical details and Chapters 6 and 7 for applications) as the tabulated values contain a part of error and may not compensate all errors as they are mainly derived from only skin roughness.

Prior roughness spatial distribution In 2D hydraulic modeling, the roughness parameter during a flood event is considered variable in space. Thus, a different value of roughness is assigned for each location in the study area. When there are no observations on the floodplains to calibrate the model, the zoning of roughness parameters typically relies on land use [Mayo et al., 2014]. The study area is divided into subdomains of homogeneous land use. For instance, the current methodology for coastal inundation in the United States is based on published land use/cover data, such as the National Land Cover Dataset. For each subdomain, a roughness value is assigned to the entire subdomain based on tables, as presented previously. The roughness values used in the subdomains for EDF flood studies (and in most of the following chapters of this work) are reported in Table 2.4, and based on CLC land use maps. CLC is a European program that provides

consistent and regularly updated information on land use across Europe every six years, with a spatial resolution below 100 meters, as visualized in Figure 2.4 for France.

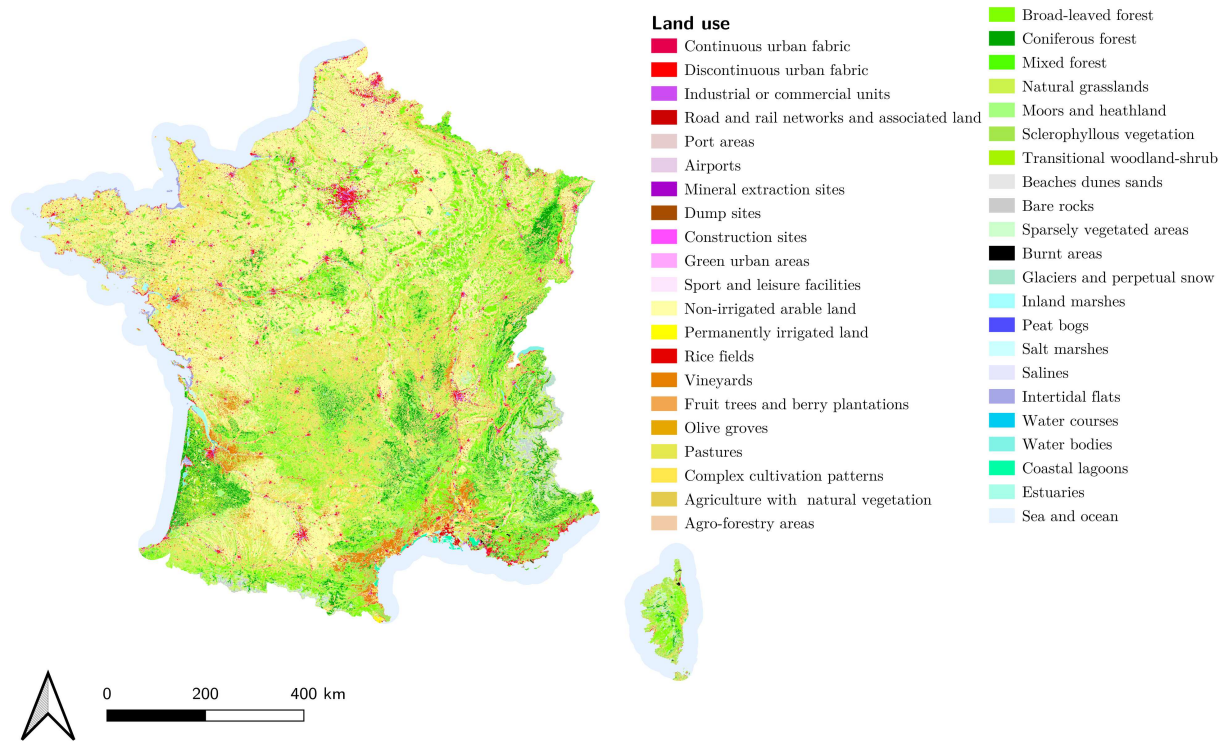


Figure 2.4: Corine Land Cover classification of land use in France.

Land use is usually classified using satellites, and many studies use satellite imagery to map it in different areas of the world (even before the CLC product) [Hossain et al., 2009; Forzieri et al., 2011; Van der Sande et al., 2003]. Using remote sensing to detect different areas leads to tremendous improvements in defining the roughness of floodplains. Straatsma and Middelkoop [2006] used airborne imagery to monitor vegetation heights, which are used to determine the distinct zones. Mason et al. [2003] advocates for defining accurately the parameters of the floodplains before any calibration to reduce the role of calibration. They resort to LiDAR to create a map of vegetation height and differentiate regions of short (<1.2 m), intermediate, and tall (>5 m) vegetation, creating subdomains of equivalent roughness.

Note: This study assumes that the floodplain roughness is divided into homogeneous land-use subdomains, as informed by existing literature and operational considerations. This approach constrains the number of parameters (i.e., number of subdomains) and the spatialization of the roughness field. During calibration, the number of parameters and spatialization are usually preserved without modification to the spatialization. Then, this spatialization into subdomains is a prior guess on the geometric structure of the roughness field. In some optimization studies, alternative methodologies use Karhunen-Loève expansions to model roughness as a stochastic process with parameters defined at each mesh node [Huang and Qin, 2014; Siripatana et al., 2018]. However, these alternatives do not benefit from prior information based on land use in each subdomain and do not capture discontinuities between subdomains which may lead to abrupt changes in roughness values (e.g., dikes, roads).

Difference in roughness between hydraulic models Most prior quantification presented in the previous paragraphs is based on land use and 1D approaches. When using this prior in 2D numerical models, the values may not represent the true flow resistance. The effective roughness depends on the model used for flood modeling, as they operate differently [Morvan et al., 2008]. Beyond the physical meaning of roughness, in hydraulic models, the roughness parameter accounts for all physical processes omitted in the modeling and compensates for discretization or local topographic errors [Morvan et al., 2008]. For instance, in 1D modeling, energy dissipation is not explicitly modeled; instead, it is often incorporated into the friction parameter. This is typically accomplished by assigning a lower Strickler value (more pronounced dissipation effects) than would be used for 2D modeling. In 1D, roughness values are assigned within the channel, and for floodplains, the values are not determined by subdomains. Instead, a single coefficient is applied to both the right and left sides of the channel for each cross-section, averaging the roughness in the transverse direction of the flow and not considering the spatial variations of roughness across the channel or floodplains. The shape of the channel is not considered in the 1D model; thus, the value between the 1D and 2D models should be modified to account for this effect, which is directly incorporated in the latter model. On the other hand, 2D models can account for flow directionality, but the roughness is often treated as isotropic (i.e., the same value in all planar directions). For a better representation of this parameter, the roughness may be considered anisotropic (varying depending on directions), due to aligned vegetation or buildings. Practically, Table 2.4 reports the differences in proposed prior Strickler values for different land uses depending on the dimension of the models being 1D or 2D based on experts' opinions, such as Chow [1959]. Even among 2D models, the friction parameter can vary, as they may not solve the same equations [Horritt and Bates, 2002].

Table 2.4: Proposed values of Strickler coefficient (K_s in $m^{1/3}s^{-1}$) for 1D and 2D models in the floodplains [Bacchi and Pavan, 2025].

Nature of the surface	In 1D	In 2D
Fields and meadows without crops	15 to 20 ⁽¹⁾	20
Cultivated fields with low vegetation	13 to 18 ⁽¹⁾	15 to 20 ⁽¹⁾
Cultivated fields with high vegetation	8 to 13 ⁽¹⁾	10 to 15 ⁽¹⁾
Shrubland and undergrowth areas	5 to 10 ⁽¹⁾	8 to 12 ⁽¹⁾
Areas of low urbanization (town)	5 to 8	8 to 10 ⁽¹⁾
Highly urbanized areas (agglomeration)	5	5 to 8 ⁽¹⁾

⁽¹⁾Depending on the density of the structures or vegetation.

2.1.3.4 Roughness parameterization: Calibration

Roughness calibration aims to adjust the value of $K_s(x, y)$ in 2D flood models to better reproduce the dynamics of the flood and achieve a better fit between the observations and the simulations. The initial estimation of the value of K_s is provided by prior estimates based on land use, as explained in 2.1.3.3. While an increasing number of studies attempt to incorporate physics to assign more precise friction parameters, it is still considered that the bottom friction should be a calibrated parameter.

Channel calibration In most studies, roughness is calibrated using data from both non-overflowing and, if available, overflowing flood events to account for the different

regimes in the channel. The main calibration data for the channel come from gauge stations that frequently measure the water depth. The calibration of the river channel with in-situ water height measurements or flood wave propagation time has been the topic of numerous articles [Pappenberger et al., 2005, 2008]. The channel reach of a hydraulic model is usually composed of heterogeneous areas, so a spatially varying friction is considered for the different parts of the reach. While a single friction parameter for the channel generally provides good results for the overall channel, dividing the channel into different roughness zones yields more locally accurate results [Ayvaz, 2013]. The channel roughness can also be calibrated with satellite water depth measurements [Domeneghetti et al., 2014]. Channel calibration can also benefit from additional information on floodplains, as the water depth on the floodplains is also sensitive to the roughness values in the channel. Schumann et al. [2007b] have used satellite images to derive roughness values for different parts of the channel based on satellite observations, thereby improving local water height estimation. It is now a widely used technique to calibrate the reach of streams based on satellite flood extent images [Aronica et al., 2002; Horritt and Bates, 2001].

Note: In this work, the focus is on the calibration of the floodplain roughness field, and the channel parameterization is based on previous calibrations using stage-gauging stations data on historical flood events. However, in Chapters 6 and 7 applications with an uncertain channel to calibrate are conducted.

Floodplain calibration Most studies consider a unique friction parameter on the floodplain [Horritt and Bates, 2002]. The choice to use a unique friction parameter arises from the lack of data and techniques to calibrate multiple subdomains based on their land-use classification. This assumption is often sufficient for small floods as it does not significantly impact modeled water depths and velocities. However, for large-scale floods, subdividing the floodplains into different areas is necessary to reproduce the flood dynamics. In Werner et al. [2005], they calibrate distributed floodplain roughness in a flood model using gauged data and flood extents. They found that, with increased floodplain roughness heterogeneity, the model's performance improves up to five floodplain parameters. Including too many parameters can lead to equifinality [Werner et al., 2005], where multiple parameter combinations produce similar model outputs [Beven, 2006]. This issue can result in compensatory effects between parameters, with outcomes that may fit the observations but lack physical realism if the parameters are not properly constrained [Pappenberger et al., 2005].

Remark: These considerations highlight the need to conduct a SA on the parameters (as described in Chapter 3) to avoid equifinality issues, and determine the maximum number of subdomains that can be calibrated with the available observations to improve the performance of the numerical model.

The calibration of floodplain subdomains is often carried out using pointwise data, such as watermarks, due to a lack of distributed spatial data [Neal et al., 2009]. Ballesteros et al. [2011] calibrated the floodplain roughness based on scars left on trees by moving riparian vegetation during floods. However, satellite images are increasingly available during flood events to calibrate the models. Different studies have extracted dense networks of water depth observations from satellite observations to calibrate the roughness [Lai and Monnier, 2009; Hostache et al., 2010; Matgen et al., 2007b; Dasgupta et al., 2021a]. Although water depth remains a key variable for constraining hydraulic models, its derivation from satellite imagery typically requires complex processing with high levels of uncertainty.

Alternatively, some studies propose directly assimilating flood extents to avoid the need for conversion into water levels [Horritt, 2000; Hall et al., 2011; Wood et al., 2016; Balbi and Lallemant, 2023]. Gobeyn et al. [2017] has shown that during the rising limb of the flood, the calibration process is more accurate because the channel’s roughness mainly influences the water’s velocity. Wood et al. [2016] also indicates that the highest information content comes from satellite images at the flood peak for the case of the Severn River in the United Kingdom.

Satellite images are now a primary source of distributed data for calibrated and validated models, as suggested by the literature. Therefore, we investigate how satellite images can enhance the quantification of roughness in floodplains for operational purposes. First, satellite observations and the type of data that can be extracted from them are introduced in the following section.

2.2 Monitoring floods with satellite observations

Earth Observation using remote sensing provides an increasing amount of data for monitoring the Earth, particularly natural hazards. For several decades, satellite missions have provided spatially distributed observations to detect, map, and analyze various natural hazards, as visualized in Figure 2.5.

For example, thermal infrared satellite observations (from MODIS mission) are used to track wildfire propagation [Wooster et al., 2013] or map burned areas [Giglio et al., 2016]. Similarly, satellite observations support the mapping of glacier displacement and volume changes [Paul et al., 2013], drought evolution [AghaKouchak et al., 2015], or meteorological modeling, where radiance, wind, and precipitation data are collected with satellites to improve the numerical models’ predictions [Bauer et al., 2015]. These examples, while not exhaustive, highlight the broad applications of satellite observations for mapping dynamical Earth systems.

Although satellites provide continuous monitoring of the Earth, revisit times between acquisitions can be important, and the observations can be difficult to interpret. Complementary airborne LiDAR and drones can help to fill these gaps. Airborne LiDAR, for example, provides high-resolution topographic data and is applied in varied applications such as canopy and biomass estimation [Cao et al., 2019], measuring post-earthquake fault displacement or damages [Nissen et al., 2012], or for snow depth estimation [Deems et al., 2013]. Drones offer rapid deployment during disaster assessments and small-scale hazard monitoring. They have been used for monitoring floods, providing real-time data during flood events [Perks et al., 2016; Karamuz et al., 2020], as well as data for reanalysis of flood events and improvement of future forecasting.

In particular, satellite observations can be used to capture images of flood events (e.g., Figure 2.5f). These images are dense observations that can be used as data for calibrating numerical flood models. Therefore, this section describes the use of satellite observations for monitoring flood events and extracting valuable data for flood model calibration. While the focus is on satellite imagery, other remote sensing technologies, such as airborne LiDAR, drones, and aerial photographs, are valuable tools for increasing data availability during a flood event [Raclot, 2003; Muhadi et al., 2020]. For further details on sensor principles and characteristics, readers can consult dedicated reviews and textbooks [Lillesand et al., 2015; Chuvieco, 2020].

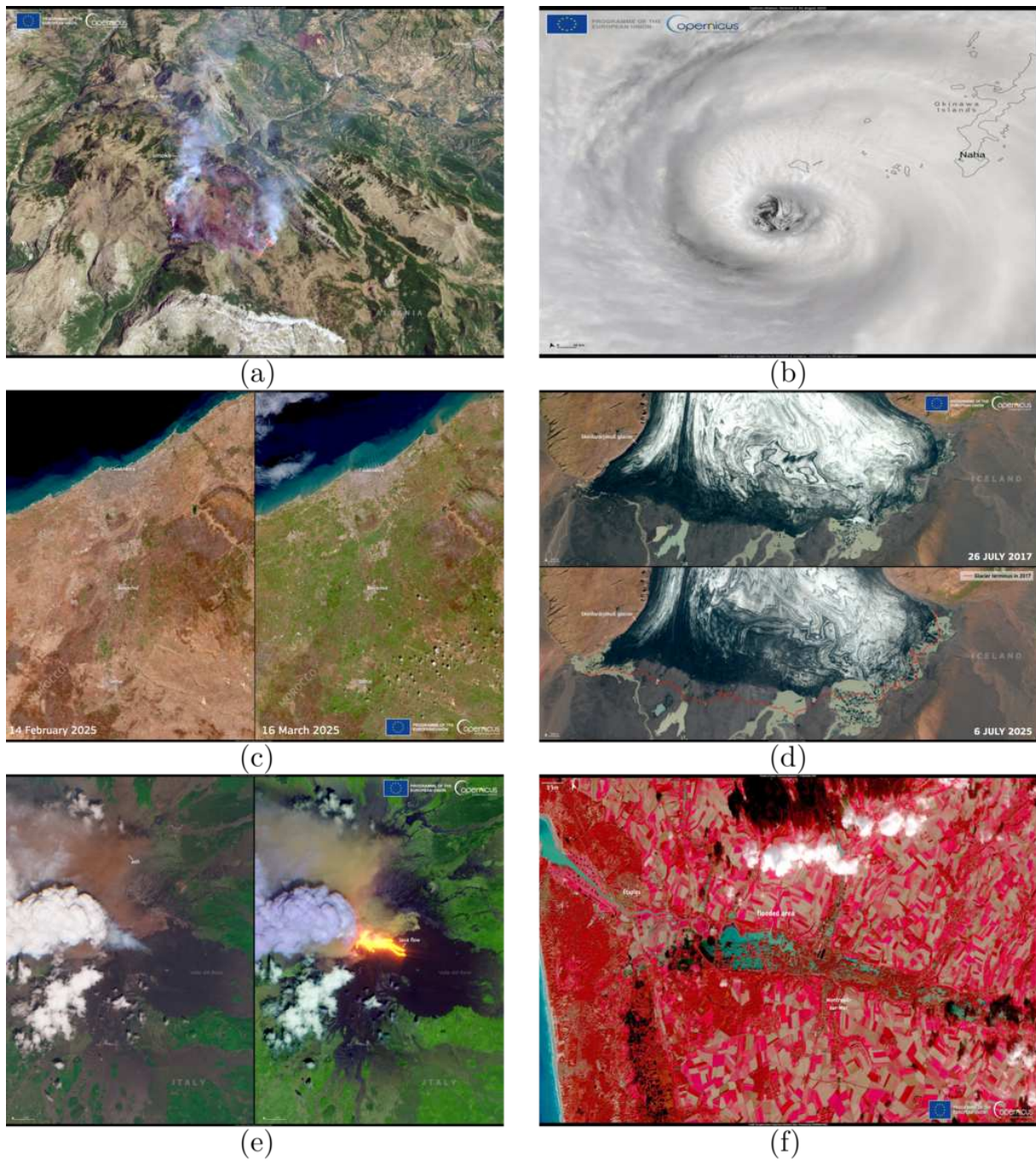


Figure 2.5: Satellite imagery of various natural hazards: (a) Wildfires in Albania (28 July 2025); (b) Typhoon Khanun impacting Japan (3 August 2023); (c) Vegetation resurgence after rainfall in Morocco (27 March 2025); (d) Skeiðarárjökull glacier, Iceland (9 July 2025); (e) Mount Etna eruption (2 June 2025); (f) Flooding after heavy rainfall in Pas-de-Calais, France (12 November 2023). Credit: European Union, Copernicus Sentinel-2 imagery.

2.2.1 Electromagnetic spectrum

Earth Observation with satellite missions relies on the emission and reception of electromagnetic radiation, which propagates through space as waves. Satellite sensors may be passive, using the Sun as an illumination source, or active, emitting their own electromagnetic radiation (Figure 2.6).



Figure 2.6: Difference between passive (left) and active (right) satellite sensors.

An electromagnetic radiation is characterized by its phase, amplitude, and wavelength ($\lambda = \frac{1}{f}$ with f the frequency). Along the electromagnetic spectrum depicted in Figure 2.7, atmospheric windows can be exploited for remote sensing applications (e.g., visible light [380 nm; 780 nm], infrared [780 nm; 0.1 mm], and microwave [0.1 mm; 1 m]) [Lillesand et al., 2015].

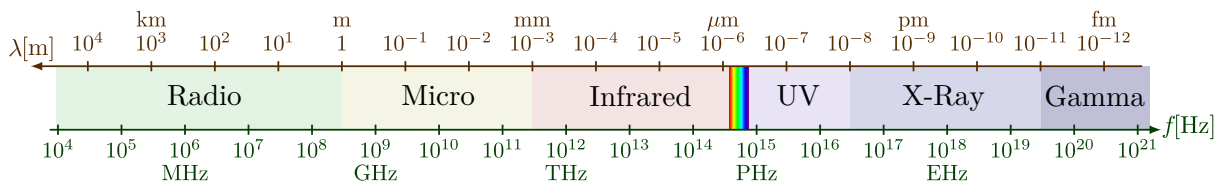


Figure 2.7: Electromagnetic spectrum with associated wavelengths (top arrow) and frequencies (bottom arrow).

2.2.2 Satellite sensors

Earth Observation satellites used for environmental monitoring can be grouped into three main categories: optical imaging satellites, radar imaging satellites, and altimeter satellites.

2.2.2.1 Multi-spectral optical sensors

Optical sensors are passive, capturing reflected sunlight during the day. Optical satellites are equipped with multi-spectral sensors to capture images in the visible and infrared portions of the electromagnetic spectrum, allowing for the detection of different features on the ground. Depending on the sensors, the spatial resolution ranges from a few centimeters to hundreds of meters. For decades, optical satellites have been the most widely used for environmental monitoring, as the images (which resemble standard photographs) in these wavelengths are easy for the human eye to analyze, as illustrated in Figures 2.5 and 2.8. Medium-resolution satellite data (e.g., Landsat and Sentinel-2, with decametric spatial resolution) have been freely available and widely used for flood applications for decades for flood mapping and extent delineation or flood damage assessment [Grimaldi et al., 2016; Schumann et al., 2022]. Thanks to this long history, different programs have been monitoring global floods with decades of free data (or through paid services for high-resolution products from commercial companies), such as the Dartmouth Flood Observatory (<https://floodobservatory.colorado.edu/>) using the widely used MODIS satellites (Terra and Aqua) with a spatial resolution of 250 m and a revisit time of twice a day, allowing the monitoring of large floods around the globe. High-resolution imaging satellites can capture small objects, but usually have low revisit times and narrow swaths.

Many optical satellites data owned by governments, non-profit organizations, or companies are available for flood applications. Table 2.5 provides a non-exhaustive list of widely used optical satellites for flood monitoring.

Table 2.5: Examples of widely used optical satellites for flood monitoring.

Mission	Years of operation	Ground resolution (m)	Revisit time (days)	Data access
Sentinel-2A; 2B	2015–; 2017–	10–60	10 (5 in constellation)	Available
Sentinel-3A; 3B	2016–; 2018–	300	4 (2 in constellation)	Available
Landsat 5	1984–2013	30	16	Available
Landsat 7	1999–	15–30	16	Available
Landsat 8	2013–	15–30	16	Available
Landsat 9	2021–	15–30	16	Available
MODIS (Aqua; Terra)	1999–; 2002–	250–1000	1	Available
Proba-V	2013–2021	300	2	Available
WorldView (1; 2; 3)	2007–; 2009–; 2014–	0.5; 0.5–2; 0.31–30	1.7; 1.1; 1	Restricted*
SPOT (5; 6; 7)	2002–2015; 2012–; 2014–	2.5–20; 1.5–6; 1.5–6 0	1–3	Restricted*
Ikonos-2	1999–2015	0.8–4	1–3	Commercial
QuickBird-2	2001–2015	0.61–2.5	1.5–2.8	Restricted*
Pleiades (1A; 1B)	2011–; 2012–	0.5	1	Restricted*

* Data access requires a research proposal via ESA (<https://earth.esa.int/eogateway>).

The main issue with optical satellites is that cloud coverage and vegetation hinder their utilization for detecting floods [Grimaldi et al., 2020]. For instance, the average annual cloud coverage in France ranges from 30% to 70% depending on the region [Direction de la Climatologie, 2009], and is especially prevalent during flood events. Furthermore, since these systems rely on recording reflected sunlight, they cannot operate at night.

2.2.2.2 Altimeters

Altimetry measures the distance between the Earth and the satellite. Altimeters were historically designed for ocean surface topography, but also measure the height of continental water on large lakes and rivers [Calmant et al., 2008; Tarpanelli et al., 2013]. A radiation signal is emitted to the ground, and the water elevation is computed by using the signal’s return time and propagation speed in the air, such as:

$$\eta = R - c \frac{\Delta t}{2} - \sum cor, \quad (2.19)$$

where η is the water elevation, R is the altitude of the satellite above a reference ellipsoid, c is the speed of light, $\frac{\Delta t}{2}$ is the travel time of the pulse, and $\sum cor$ accounts for ionospheric, tropospheric, and tidal corrections. The uncertainty in water height usually ranges from 1 to 50 cm (the computed water level is the mean over the section on a large footprint, spanning hundreds of meters to kilometers). Altimeters like Envisat, Topex/Poseidon, and Jason provide point-based water levels, with high accuracy for lakes (<10 cm) and moderate accuracy for rivers (20–50 cm) [Crétaux and Birkett, 2006; Alsdorf et al., 2007b].

Table 2.6 provides an overview of altimeters used for measuring continental water surfaces. In recent years, databases have been established to compile altimetry-based time series of water depth over rivers and lakes, including available points and their range of uncertainty, such as Hydroweb (<https://hydroweb.theia-land.fr/>) or the Database for Hydrological Time Series of Inland Waters (<https://dahiti.dgfi.tum.de/en/products/water-level-altimetry/>). For flood calibration, these altimeters can be used for large rivers; however, the along-track spacing varies from 500 m to several kilometers. With the launch of the SWOT swath altimetry satellite, it will become feasible to measure spatial fields of surface water elevation with a high spatial resolution. The accuracy is expected to be around 10 cm over 1 km² and 25 cm over 250 m².

Table 2.6: Examples of widely used altimeters for water level measurements.

Mission	Years of operation	Revisit time (days)	Inter-track distance (km)	Data access
SWOT	2022–	21	Swath altimetry	Available
ERS-1 RA	1991–2000	3/35/168	80 (35 days)	Restrained*
ERS-2 RA	1995–2011	35	80	Restrained*
ENVISAT RA-2	2002–2012	35	80	Restrained*
JASON-1/2/3	2001–2013; 2008–; 2016–	10	315	Available
Sentinel-3A/3B	2016–; 2018–	27	104; 52	Available
TOPEX/Poseidon	1992–2006	10	157	Available

* Data access requires a research proposal via ESA (<https://earth.esa.int/eogateway>).

2.2.2.3 Synthetic Aperture Radar

SAR sensors are active sensors independent of external illumination sources, emitting waves in the microwave portion of the electromagnetic spectrum. They guarantee continuous observation with an active system that emits microwave pulses and penetrates cloud coverage, operating day and night, unlike optical sensors [Smith, 1997]. The ratio between the emitted and received signals by the satellite antenna depends strongly on the roughness properties of the surface from which it is reflected. Open water surfaces are smooth and reflect a large portion of the incoming radiation away from the sensors [Ulaby, 1982], which results in low reflected energy, leading to “dark spots” on the satellite image. Dry areas, on the contrary, are rougher and reflect a larger proportion of the radiation back to the antenna and therefore appear brighter on the images, as visualized in Figure 2.8.

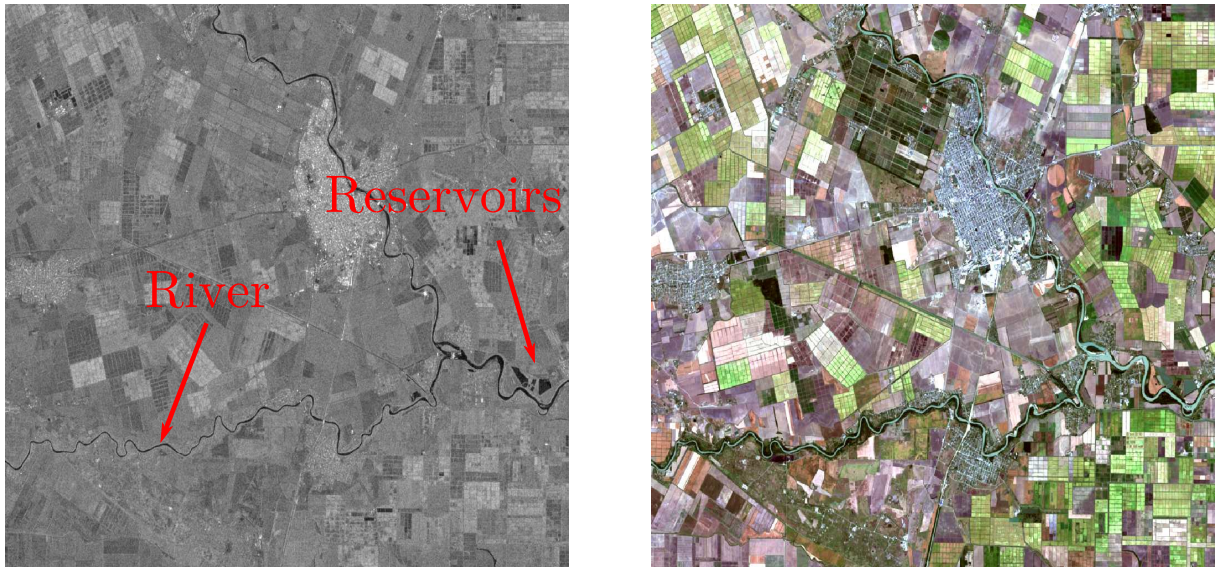


Figure 2.8: Comparison of a Synthetic Aperture Radar Sentinel-1 image (left) with an optical Sentinel-2 image (right) acquired at the same time.

For more than 30 years, SAR satellites have been widely used for detecting floods from space [Schumann et al., 2009; Grimaldi et al., 2016]. Table 2.7 reports some of the most used SAR satellites for flood assessment.

Table 2.7: Examples of widely used SAR satellites for flood applications.

Mission	Years of operation	Ground resolution (m)	Revisit time (days)	Polarization	Band	Data access
Sentinel-1A; 1B	2014-; 2016-	5–100	12 (6 in constellation)	Dual	C	Available
COSMO-SkyMed	2007-	15–100	16 (4 in constellation)	Single or Dual	X	Restrained*
TerraSAR-X	2007-	1–16	11	Full	X	Commercial
TANDEM-X	2010-	3	11	Full	X	Commercial
RADARSAT-1	1995–2013	8–100	24	HH (single)	C	Restrained*
RADARSAT-2	2007-	3–100	24	Full	C	Restrained*
ERS-1	1991–2000	25	35	VV (single)	C	Restrained*
ERS-2	1995–2011	25	35	VV (single)	C	Restrained*
ENVISAT-ASAR	2002–2012	30–1000	35	Single or Dual	C	Restrained*
ALOS-PALSAR	2006–2012	10–100	46	Single or Dual	L	Available
ALOS-PALSAR-2	2014-	3–100	14	Single or Dual	L	Commercial
JERS-1	1992–1998	18	44	HH (single)	L	Restrained*
SAOCOM-1; 2	2018-; 2020-	10–100	16 (8 in constellation)	Full	L	Available
Kompsat-5	2013	1–20	28	Full	X	Commercial
RCM	2019-	3–100	12 (4 in constellation)	Full	C	Commercial
PAZ	2018-	1–15	11	Full	X	Commercial
SeaSat	1978	25	/	L	L	Available

* Data access requires a research proposal via ESA (<https://earth.esa.int/eogateway>).

Note: In this research, we focus on SAR for observing flood events. In the literature, it is the most commonly used type of satellite observation for monitoring floods, thanks to its ability to see through clouds and at night. For example, [Tarpanelli et al. \[2022\]](#) demonstrates, through an analysis of Sentinel-1 optical satellite data, that for two satellite constellations orbiting in both ascending and descending orbits, Sentinel-1 which can detect floods even with clouds could detect 58% of flood events in Europe (limited by the revisit time of the satellite in comparison to the short duration time of most flood events). When using optical images, obstructed by cloud cover, only 28% of floods can be detected.

While clouds are often present in optical satellite images, these data can be used when they are cloud-free or to retrieve partial information on the study area when the cloud cover is not complete. The calibration process of the flood model using flood maps (Chapter 6) or water depth fields (Chapter 7) extracted from SAR imagery could also be applied with observations extracted from optical images.

To our knowledge, floodplain roughness has not yet been calibrated using point-wise data from altimeters (excluding swath altimetry), due to the sparse spatial coverage and high revisit time. They have only been used for calibrating channel roughness [[Domeneghetti et al., 2014, 2021](#)]. However, in the operational context, while not sufficient for calibrating floodplain roughness, they can still be used as validation data and as additional points with low uncertainty in the calibration process.

SAR images are more challenging to interpret because the backscatter intensity is linked to the roughness and dielectric properties of the images. However, they present a valuable opportunity for flood monitoring. These properties are described in more detail in Section 2.2.3.

2.2.3 A focus on Synthetic Aperture Radar

SAR observations appear as a useful source of distributed data for flood monitoring. Their characteristics, such as spatial resolution, radiometric effects, geometric distortions, scattering mechanisms, or the interaction of the SAR signal with water surfaces, must be detailed to understand how to treat the images to extract data for calibrating flood models and the sources of uncertainties.

2.2.3.1 Principle of radar imaging

Radar sensors can be both imaging and non-imaging. Imaging sensors provide a 2D image of the Earth's surface, while non-imaging sensors (including altimeters) compute only the travel time of the emitted wave. We focus on imaging radar sensors with lateral viewing. Radar sensors are based on an antenna that emits electromagnetic pulses of duration τ (microwaves to radio waves), in the LOS direction as illustrated in Figure 2.9. When the signal touches the ground, a portion is reflected (backscattered) back to the antenna, where its amplitude and phase are recorded (mono-static sensor). The received signal is sampled according to its arrival time. The start of the received signal corresponds to the part of the signal with a shorter travel time that interacted with the closest pixels to the antenna (near range), while the end of the reception corresponds to the farthest pixels (far range). The distance between the closer and farther illuminated pixels in the LOS direction is referred to as the swath. The distance between the antenna and the ground, denoted as slant range R , can be calculated. The nadir direction is the direction pointing straight down from the satellite to the Earth's surface. The look angle α of the LOS direction is defined relative to this nadir. The incidence angle between the radar signal and the normal to the Earth's surface is α_i with differences between α and α_i due to the curvature of the Earth [Lillesand et al., 2015]. All the notations are summarized in Figure 2.9.

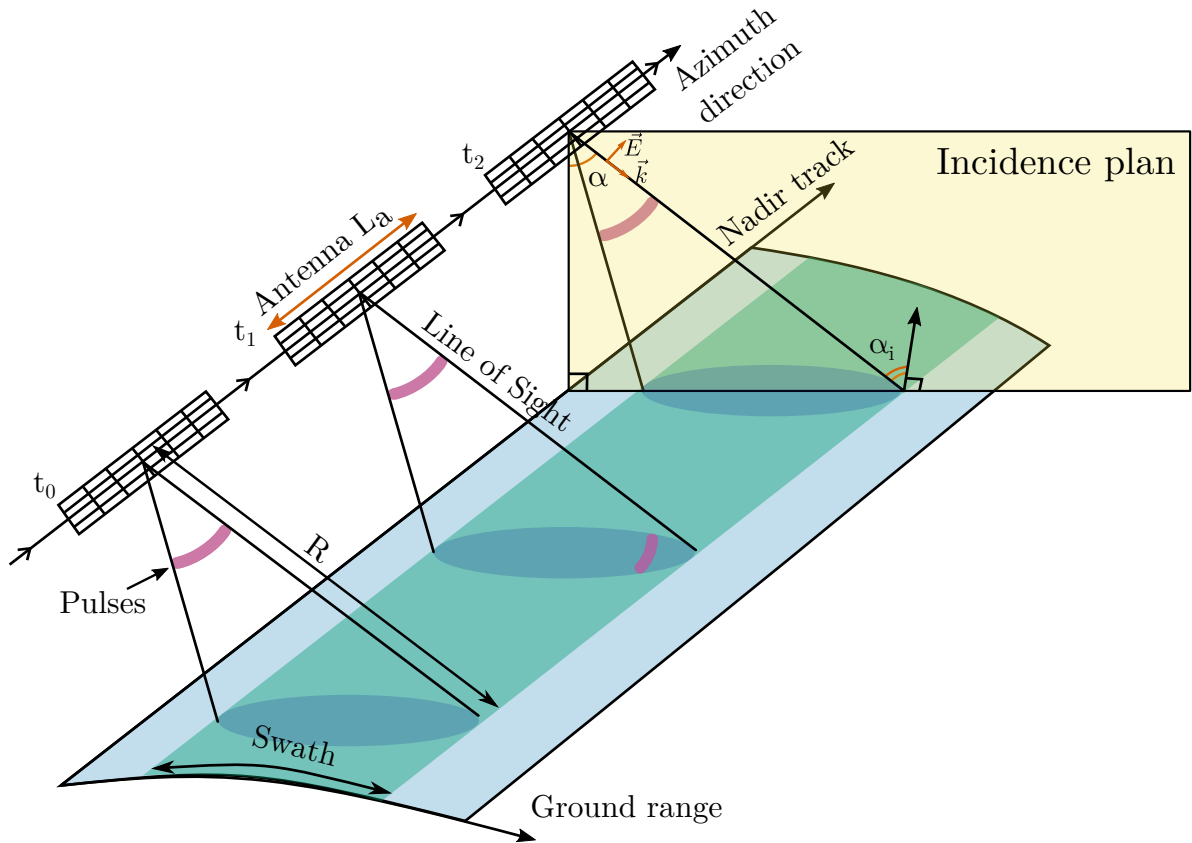


Figure 2.9: Acquisition geometry of a side-looking radar imager. The radar moves along the azimuth direction and emits pulses in the Line Of Sight direction.

Spatial resolution The spatial resolution of a radar image is the shortest distance that the sensor can distinguish between two objects. The resolution varies between the azimuth and LOS directions. In the LOS direction, the resolution is a function of the

duration of the emitted pulse τ , which is modulated in frequency (ensuring pixels are illuminated with a different frequency and a sufficient power). Two illuminated objects at the same time cannot be distinguished, so τ should be small.

Therefore, the spatial resolution of a pixel in the LOS direction in the sensor geometry is:

$$r_s = \frac{c\tau}{2} = \frac{c}{2B}, \quad (2.20)$$

where c is the celerity of light in air, and B is the bandwidth of the pulse. The spatial resolution can be converted into the ground-range resolution using the incidence angle, such as:

$$r_G = \frac{r_s}{\sin(\alpha_i)}. \quad (2.21)$$

For the azimuth direction, the resolution depends on the antenna's beamwidth β :

$$r_A = R\beta = R\frac{\lambda}{L_A}, \quad (2.22)$$

where L_A is the length of the antenna and λ is the wavelength. Thus, to have a high-resolution the size of the antenna L_A should be important. The azimuth and LOS ground spatial resolution depends on the wavelength and incidence angle, so across the scene, the spatial resolution differs (e.g., the resolution improves when the pulse length τ decreases). In practice, for a ground resolution of 10 m with a wavelength of 5.6 cm, an incidence angle of 35° , and an altitude of 700 km (Sentinel-1 characteristics), an antenna of over 7 km is required. To address this issue, SAR sensors were developed to create a synthetic antenna of that length using a shorter antenna.

Principle of Synthetic Aperture Radar SAR refers to radar systems that replicate an extended antenna using the Doppler effect with a collection of radar echoes from a hypothetical stationary ground scene as the system moves along the azimuth direction. Because of the Doppler effect, the front part of the beam shifts, while the back part shifts inversely. Each area on the ground is detected multiple times (depending on the beam angle β), and their Doppler shifts are processed to produce a synthesized acquisition from a single sensor with a longer antenna, as illustrated in Figure 2.10. Pixels in the far range are observed more often than those in the near range. Finally, the resolution can be deduced from the actual antenna length as [Ulaby et al., 1986]:

$$r_{A,SAR} = \frac{L_A}{2}. \quad (2.23)$$

2.2.3.2 SAR backscatter

The wavelength is small compared to the length of propagation of the waves. Therefore, the waves are approximated as a plane progressive wave. The electric field of a planar progressive wave is:

$$\vec{E}(\vec{r}, t) = \begin{cases} E_x = a_x \cos(\omega t - \vec{k} \cdot \vec{r} + \delta_x) \\ E_y = a_y \cos(\omega t - \vec{k} \cdot \vec{r} + \delta_y) \\ E_z = 0, \end{cases} \quad (2.24)$$

where $\vec{r}$ is the position vector, $||\vec{k}|| = \frac{2\pi}{\lambda}$ is the wave number with λ the wavelength, a_x and a_y are the amplitudes, $\omega = c||\vec{k}||$ is the angular frequency, and δ_x and δ_y are the initial phases of the waves.

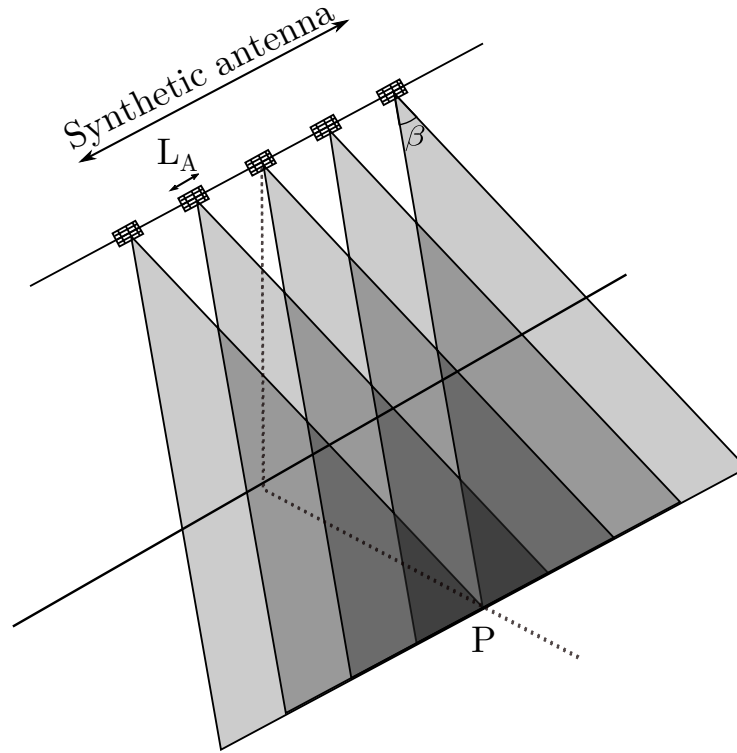


Figure 2.10: Principle of a Synthetic Aperture Radar sensor capturing the same point P multiple times with an antenna of length L_A to create a “synthetic” antenna using the Doppler effect. The shades of grey show the number of times the same area is seen in the swath.

The ratio between the emitted wave and the reflected wave can be computed since the emitted wave is known and the received wave is measured. More generally, assuming that the energy is reflected isotropically [Woodhouse, 2017], the RCS writes:

$$\sigma = 4\pi R^2 \frac{|\vec{E}_r|^2}{|\vec{E}_i|^2}. \quad (2.25)$$

The RCS increases when the area of measurement increases (variable in the scene), so to compare the coefficients, the RCS is normalized by the geometrical area of the ground surface A to obtain the backscattering coefficient [Woodhouse, 2017]:

$$\sigma^0 = \frac{\sigma}{A}. \quad (2.26)$$

It should be noted that the power received back to the sensor not only depends on the RCS of the target but also on the sensor parameters (slant range distance R and the gain of the antenna G). For more details on a more comprehensive formulation that includes both G and R , the reader can refer to Ulaby et al. [2015]. As such, a 2D satellite image can be constructed with each pixel’s backscattering coefficient depending on the surface and sensor parameters.

Note: In the following, we focus our attention on SAR intensity imagery only based on the backscattering coefficient. However, it is also possible to use the phase of the signal for InSAR or the polarization for SAR polarimetry.

2.2.3.3 Radiometric noise

SAR images may contain granular noise texture called speckle (speckle pattern or speckle noise) [Goodman, 1976], which looks like salt-and-pepper, with some very dark pixels and some very bright pixels, as visualized in Figure 2.11.

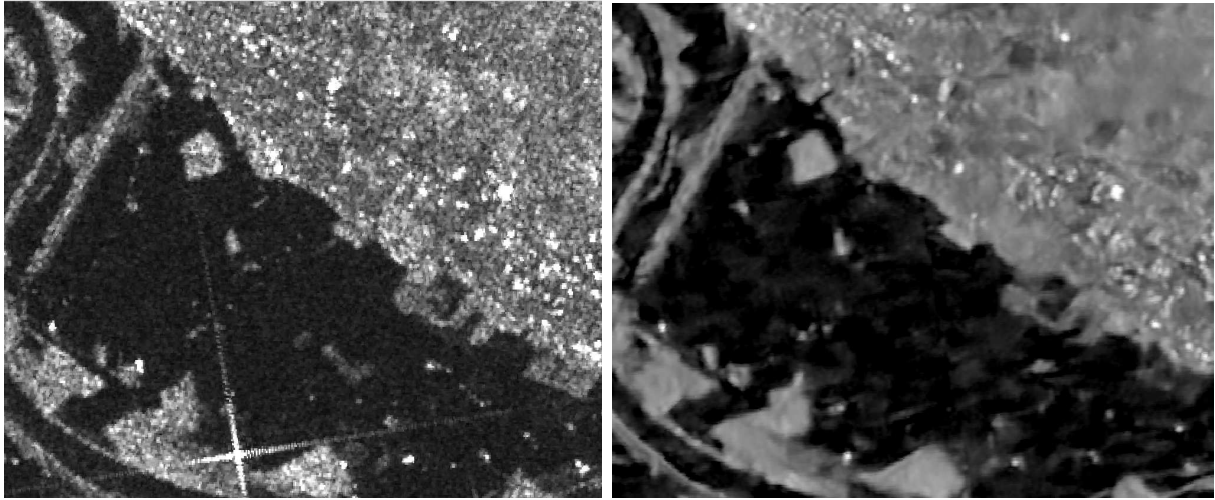


Figure 2.11: Illustration of speckle noise on a Sentinel-1 image on the left and without speckle on the right. The black zones correspond to water, and the brighter zones to pasture areas.

SAR sensors emit only a single wavelength, so the waves interfere with each other, creating interference. Random constructive and destructive interferences are generated from many out-of-phase scattering echoes within a resolution cell [Gagnon and Jouan, 1997]. Speckle hinders the interpretation of SAR images, making it difficult to determine the actual value of the backscatter. Pre-processing aims to remove speckle noise while preserving the contours and ideally maintaining the spatio-temporal resolution of the image. Speckle is a physical phenomenon; therefore, temporal and spatial models of the speckle statistics are built to reduce speckle noise and estimate the underlying backscatter.

Spatial filtering Spatial filtering is performed on a single observation to reduce the speckle effect, albeit at the cost of spatial resolution. A review of spatial filtering techniques is available in Deledalle et al. [2014], or Lee and Pottier [2017]. The earliest filters averaged the mean or median value of a defined rectangular window around a pixel [Huang et al., 1979]. While averaging is reasonable in homogeneous land use areas where the backscattering is expected to be similar, this approach is too simplistic and can remove the edges of features. These basic implementations have been enhanced in adaptive filters, which use local statistics of the images within small windows around each pixel [Lee, 1980; Frost et al., 1982; Lee et al., 2008].

In recent years, new types of filters have been developed with more advanced treatments, such as non-local mean filters that search for neighboring similar pixels [Di Martino et al., 2016] or methods that take advantage of the phase or polarimetric properties from different channels of SAR images and not only the backscattering intensity [Deledalle et al., 2017]. With the growing number of observations, data-driven approaches are developed to infer the backscatter intensity of satellite images based on non-supervised or supervised training on datasets [Dalsasso et al., 2021a,b]. Each type of filter has its advantages; thus, they are often combined to enhance the results of satellite image pre-processing. The main drawback of spatial speckle filtering is that it smooths the images, potentially resulting in a loss of information by artificially lowering the image resolution.

Temporal filtering Temporal filtering uses several observations of the same orbit and with the same incidence angle to reduce speckle. If there is a temporal dynamic (e.g., a flood event) with changing targets between the temporal observations, this information is lost or smoothed. However, these techniques efficiently reduce speckle and can be applied to study static systems (e.g., land use evolution over the years). Basic methods of temporal filtering apply a mean or median filter to a temporal series. This straightforward approach has been used for a long time, but similar to spatial filtering, data-driven approaches are emerging for multi-temporal filtering of speckle noise [Meraoumia et al., 2023].

Note: As discussed, the raw SAR image are subject to speckle noise. This noise hinders the interpretation of the imagery and limits the capability of the methods presented in Section 2.2.5 to delineate the flood extent. In most flood studies using SAR imagery, a single speckle noise filtering is used without taking into account the uncertainties due to the preprocessing in the generated observations. In Chapter 4, the impact of these filters on the generation of flood maps and water depth fields is evaluated.

2.2.3.4 Geometric effects

SAR sensors emit waves at an oblique incident angle, which causes geometric distortions in the images. Terrain-related distortions commonly include radar shadowing, layover, and foreshortening. These effects especially occur in mountainous regions or urban environments and are illustrated in Figure 2.12 and are described hereafter:

- Foreshortening occurs when slopes facing the sensor appear shorter in the radar image (slant-range) than in the ground range, as if they are 'leaning' towards the sensor (e.g., in Figure 2.12, $|A'B'| < |AB|$). This phenomenon intensifies for lower incidence angles as one approaches the nadir direction.
- Layover occurs when the tops of objects, such as mountains, are imaged ahead of their bases. The top of the object is closer to the nadir than its base, making the object seem to "lay over" its lower part.
- Shadows appear when the slopes facing away from the antenna are steeper than the complementary incidence angle at the top of the slope. In these regions, the image is dark due to the absence of a measured signal. Then, the area behind the mountains is unseen. The phenomenon becomes more pronounced as the incident angle increases or the local slope becomes steeper.

Other types of distortion, such as those caused by the motion of the satellite or moving targets on the ground (e.g., cars or ships), may also be present in SAR images. The interested reader can refer to Ulaby et al. [2015] and Woodhouse [2017] for further details.

2.2.4 Interaction of SAR with the Earth's surface

When the waves hit the ground, part of the signal is reflected to the antenna. To evaluate the backscatter intensity (Equation (2.26)), the power received by the antenna's sensor is measured. This quantity mainly depends on the interaction between the signal and the surface, as well as the properties of the SAR system. The predominant mechanisms impacting the energy received by the antenna are described in the following subsections.

resulting in specular reflection. The signal is directed in the opposite direction (the angle of incidence equals the angle of reflection). In contrast, for rough surfaces, the signal is reflected in various directions, a process referred to as diffuse scattering.

- Volume scattering refers to the scattering of signals in a medium or volume containing discrete scatterers with different directional responses (e.g., snow, vegetation, or leaves and branches in trees). The signal scatters multiple times within the volume before returning to the sensor.
- Double-bounce scattering occurs in urban areas due to the presence of buildings, which create corner reflectors. The signal is then redirected to the antenna due to the double bounce on the smooth surfaces.

Remark: For water surfaces and floods, specular reflection is the main scattering mechanism. Therefore, little energy is transmitted back to the antenna, resulting in a low backscattering intensity in these areas. This is the main property used to identify flooded areas on SAR imagery. More details on the detection of flooded areas are given in Section 2.2.5.

2.2.4.2 Influence of wavelength

The wavelength of the waves is crucial as the scattering mechanism depends on it, as discussed in the previous subsection. Furthermore, the penetration capacity of the signal depends on the wavelength. For longer wavelengths, the signal generally penetrates more deeply into the medium. Depending on the application, the wavelength of the satellite sensor is selected differently, as described in Table 2.8. For example, in vegetated areas, X-band SAR sensors mostly scatter at the tops of tree canopies, while C-band and L-band signals penetrate deeper into the vegetation volumes, as visualized in Figure 2.15. Similarly, in bare surfaces such as alluvial soils or glacier ice, X-band signals scatter close to the surface, while C-band and L-band signals penetrate deeper into the medium. For flood mapping, the usual wavelength is below 10 cm in the C-band. For instance, for Sentinel-1, the wavelength is around 5.6 cm.

Table 2.8: Typically used wavelengths for SAR applications extracted from Flores-Anderson et al. [2019].

Band	Wavelength (cm)	Typical applications
Ku	1.7–2.4	Mainly used for satellite altimetry
X	2.4–3.8	High spatial resolution for urban monitoring, ice and snow detection, with little penetration into vegetation cover
C	3.8–7.5	Most used band, for global flood mapping, change detection, and monitoring areas with low to moderate vegetation
S	7.5–15	Agricultural monitoring; extends C-band for denser vegetation
L	15–30	Medium resolution for geophysical monitoring, biomass and vegetation mapping, with high penetration
P	30–100	Biomass and vegetation mapping

2.2.4.3 Influence of incidence angle

The surface roughness is perceived differently depending on the incidence angle, as visualized in Figure 2.15. The backscatter can vary across the swath, not because of any

physical differences, but only due to the incidence angle. Furthermore, between the near range and the far range, the pixel size changes because of variations in the incidence angle. In the near range, the scene is often much brighter than in the far range of the swath due to stronger backscattering at lower incidence angles. Consequently, radiometric calibration and incidence angle normalization are often required to compare the backscatter across the entire swath.

2.2.4.4 Target properties

The interaction of the signal strongly depends on the physical properties (e.g., local variations in height and roughness) and electrical properties of the target.

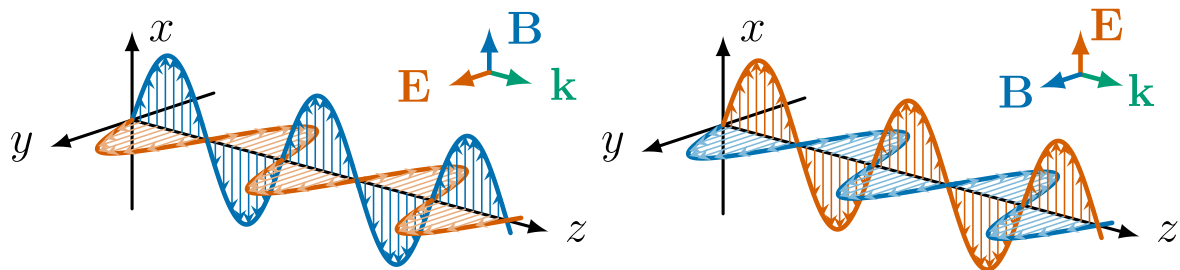
One key parameter is the dielectric constant (or relative permittivity), which measures how easily electromagnetic waves can propagate through or interact with a medium. Mediums with low dielectric constants (e.g., sand, ice) allow waves to pass through with minimal interaction, while those with high dielectric constants reflect or absorb the signal. Water has a high dielectric constant compared to other media (more than 30, while other dry media have a dielectric constant lower than 5); therefore, soil moisture is one of the most influential factors on the dielectric constant of a medium. When propagating through a dielectric medium, the waves attenuate and lose energy with exponential decay. For low dielectric constants, such as those found in sand, penetration can reach several meters. In contrast, in wet or highly conductive media, it is limited to the surface layer.

The scattering mechanisms also depend on surface roughness. However, roughness is a relative quantity of the wavelength size, as visualized in Figure 2.15. An object may appear smooth with a longer wavelength, while the same surface may be seen as rough with a shorter wavelength. According to the Rayleigh criterion, a surface is smooth if:

$$\sigma_h < \frac{\lambda}{8\cos(\alpha_i)}, \quad (2.27)$$

where σ_h is the standard deviation of local height variations of the target (considering the mean height of a target and its deviations), λ is the wavelength, and α_i is the incidence angle.

2.2.4.5 Influence of polarization



(a) Electric field perpendicular to the incidence plane: Horizontal polarization.

(b) Electric field parallel to the incidence plane: Vertical polarization.

Figure 2.14: Polarization of the electric field of a planar wave.

Polarization of an electromagnetic wave refers to the orientation of the plane of oscillation in the orthogonal plane to the wave propagation, i.e., horizontal (H) or vertical (V) planes. For a linear wave, the polarization is visualized in Figure 2.14 in the incident

plane (xOz). Four combinations of polarizations are possible: HH , VV , VH , and HV , with the first letter referring to the emitted signal and the second letter indicating the received signal. Depending on the land use, the polarization influences the measured energy and hence the backscattering coefficient. According to the literature, HH polarization is preferred for flood detection, because it is less sensitive to small roughness changes in the water surface compared to VV polarization [Henry et al., 2006].

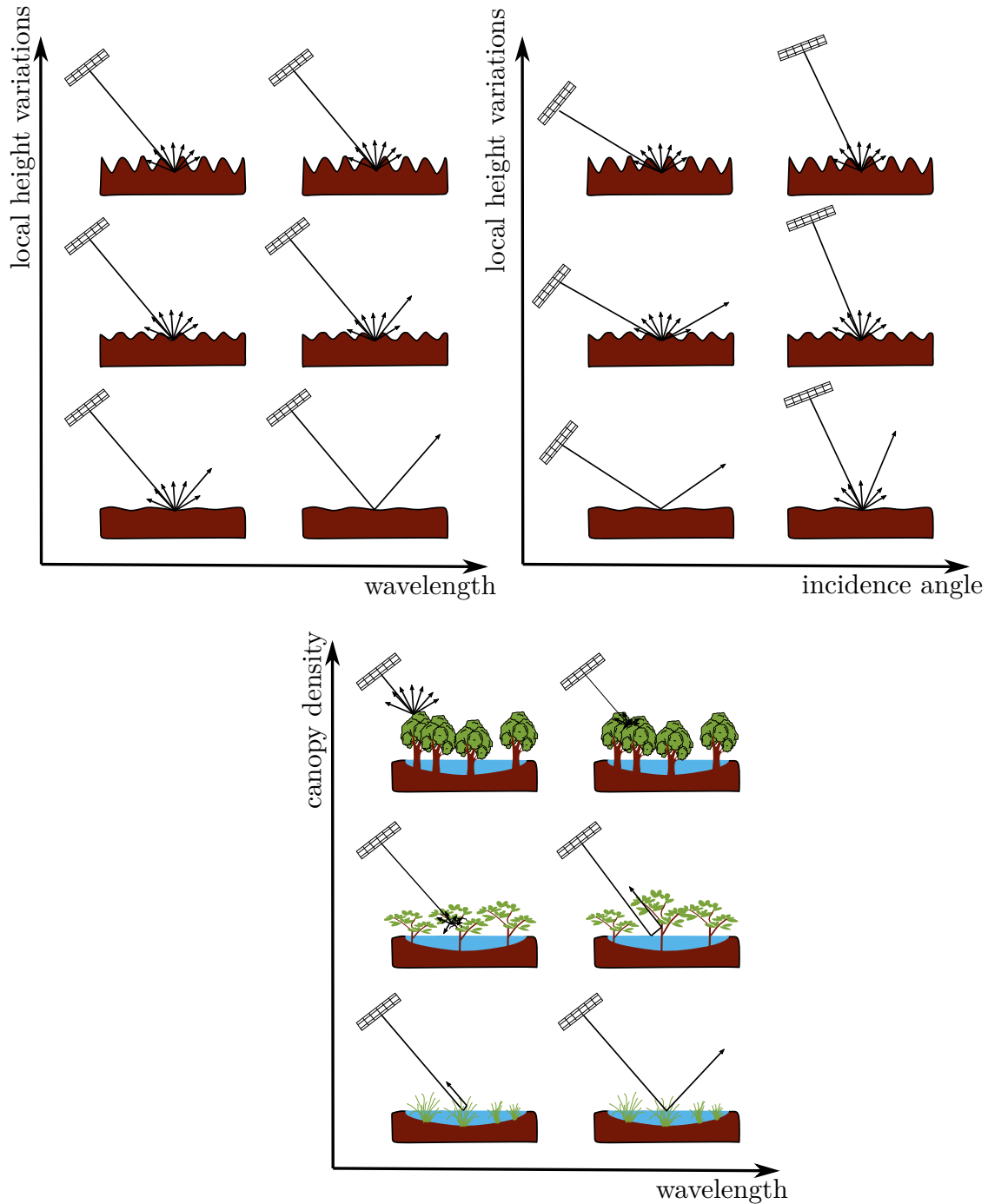


Figure 2.15: Influence of local height variations, incidence angle, wavelength, and canopy density on backscatter mechanisms adapted from Lillesand et al. [2015] and Landuyt [2021].

2.2.5 Detection of water surfaces with SAR observations

Thanks to the properties of SAR sensors, water surfaces, and notably floods, can be identified on the images. Figure 2.16 summarizes the application of SAR images for flood detection. A brief overview of the detection of water surfaces in open water, vegetated areas, and urban areas is provided, as well as how data are extracted from images in the existing literature.



Figure 2.16: Mind map of flood mapping with Synthetic Aperture Radar observations.

2.2.5.1 Open water surfaces

Because of specular reflection of water surfaces and high permittivity, open water surfaces are easily identified on SAR images, appearing as dark homogeneous regions compared to the brighter surroundings of dry areas [Ulaby, 1982]. SAR sensors depict a high sensitivity to surface water, but they can misclassify water bodies due to the confusion between land surface areas and water, which may have similar backscattering coefficients (e.g., dry

sand, concrete areas, roads, parking lots, plane tarmacs) [O’Grady et al., 2011; Martinis et al., 2018]. Strong winds or rain can increase the roughness of the free surface of water, causing more diffuse reflection. For open-water surfaces, classical approaches for segmenting flood extent based on a single image or a stack of images are described in the following paragraphs.

Visual interpretation Visual interpretation was the most widely used technique at the beginning of SAR mapping for floods, with experts manually digitizing flood boundaries based on backscattering intensity, flood dynamics, and visible terrain features, such as slope breaks, on topographic maps [Oberstadler et al., 1997; Brakenridge et al., 1998]. This method provides a preliminary estimate, but it is subjective, labor-intensive, and not automated, making reproduction difficult; the results depend heavily on expert knowledge [Mason et al., 2012a].

Thresholding Thresholding on the backscatter intensity is the most straightforward technique for automated flood mapping. The dark pixels, corresponding to flooded areas, are separated from the brighter pixels, which represent land. Automatic thresholding methods are commonly used to determine optimal thresholds [Sezgin and Sankur, 2004], such as those proposed by Otsu [1979], Kittler and Illingworth [1986], and Yen et al. [1995]. While global thresholding on the whole image is simple, the separability between classes strongly depends on the bimodality of the histogram of backscattering intensity. Classification errors may occur due to wind roughening or in the presence of bare soil, so Hostache et al. [2009] proposed a dual-threshold approach, improving edge detection by classifying pixels as certainly, potentially, or not flooded instead of using two classes. To address spatial variability, local thresholding approaches have been suggested [Bovolo and Bruzzone, 2007]. Local thresholding divides the image into representative areas where both classes can be dichotomized before threshold selection [Martinis et al., 2009, 2015b; Twele et al., 2016; Chini et al., 2017].

Active contour and region growing Thresholding methods often yield noisy classifications due to a lack of texture information. Active contour models, or snakes, detect flood boundaries by minimizing an energy function based on texture and pixel statistics [Horritt, 1999; Chan and Vese, 2001]. The curve is usually initialized using thresholding approaches [Heremans et al., 2003]. Pixels within the contour that match flood-like statistics are classified as flooded [Horritt et al., 2001]. Region growing expands an initial flooded region and then checks if the neighboring pixels satisfy a condition (usually homogeneous statistics with neighbors) to be added to the delineated flooded region [Matgen et al., 2011; Giustarini et al., 2012]. This step is repeated until no more pixels can be added to the flooded region. Region growing is often used alongside other flood mapping methods to refine their outcomes [Martinis et al., 2015a; Matgen et al., 2011].

Change detection Change detection relies on SAR time series instead of a single image. This technique aims to compare the same area at different observation times [Ban and Yousif, 2012]. With reference to a non-flooded image, flooded pixels can be detected through a decrease in the backscattering coefficient. Early techniques used only pairs of images due to limited data [Rignot and Van Zyl, 1993; Bazi et al., 2005]. However, when comparing images, the choice of reference images is essential [Hostache et al., 2012; Schlaffer et al., 2015; Li et al., 2018]. For instance, vegetation may change with seasonality and could be misclassified as water with a change in backscattering [Zhao et al., 2021]. Multi-temporal analysis using image stacks (series of observations) helps exclude

consistently low-backscatter areas, such as sand and vegetation, or assess seasonal effects [Clement et al., 2018; Schlaffer et al., 2015]. Similarly, a threshold should be specified to distinguish flooded from land pixels when comparing a series of observations.

Supervised classification In recent years, supervised methods have been increasingly used to detect floods in SAR images [Bentivoglio et al., 2022]. Models are trained on labeled datasets to identify classes in other images. Supervised methods became more feasible with the release of labeled datasets like Sen1Floods11 [Bonafilia et al., 2020] and WorldFloods [Mateo-Garcia et al., 2021]. The supervised approaches used methods such as RF [Breiman, 2001], SVM [Vapnik, 1999], and KNN [Fix and Hodges, 1951], demonstrating high accuracy ($>90\%$) across various test cases [Thanh Noi and Kappas, 2017]. CNN have seen rapid development in recent years, outperforming traditional image processing techniques [Wieland and Martinis, 2019; Bonafilia et al., 2020].

Probabilistic mapping Probabilistic flood mapping estimates the likelihood of each pixel being flooded, unlike traditional deterministic flood maps, which dichotomize the pixels into two classes. Bayesian approaches are used to generate flood probability maps by estimating the posterior probability of flooding for each pixel based on its backscatter value [Westerhoff et al., 2013; Giustarini et al., 2016]. These methods are enhanced by incorporating additional information such as topography, distance from water bodies, and InSAR coherence [Refice et al., 2014; D’Addabbo et al., 2016]. Another common approach for estimating the uncertainty of a pixel state is based on fuzzy set theory. Instead of discrete class labels, fuzzy classification assigns degrees of membership to each pixel across different states. Dasgupta et al. [2018] used fuzzy classification to derive possibility maps that represent the spatial uncertainty of flood extent.

Note: All these methodologies for extracting flood maps from the backscattering coefficient of SAR imagery have been tested on various rivers in the literature. Schumann et al. [2009] and Di Baldassarre et al. [2011] highlight that the different methods for flood mapping have their advantages and limitations, but no technique can be identified as optimal in every case. However, operational studies often select a flood mapping strategy without comparing the generated flood maps with those of other strategies, so further investigation is still required, especially to characterize the uncertainties associated with the choice of flood mapping method. Few works evaluate the uncertainties due to the choice of flood mapping method and its hyperparameters [Landuyt et al., 2018]. Therefore, in Chapter 4, these flood mapping methods (e.g., thresholding, change detection, active contour, and supervised classification) are tested on a study area with various hyperparameter configurations and using different speckle reduction methods. Indeed, the results of the flood mapping methods may differ depending on the preprocessing applied to the SAR images.

2.2.5.2 Vegetated areas

Detecting water in vegetated areas is more challenging than in open areas, due to the complex interactions of the signal with the canopy. While the dominant scattering mechanism for open water surfaces is specular reflection, flooded vegetated areas display a wider range of scattering mechanisms. These mechanisms mainly depend on the capability of the signal to penetrate through the canopy, depending on the wavelength and structural characteristics of the vegetation, as shown in Figure 2.15. When the signal successfully penetrates through the canopy and reaches water surfaces, it often results in

double-bounce scattering with vertical structures (e.g., stems, leaves, branches), leading to an increase in backscatter intensity compared to non-flooded conditions [Pulvirenti et al., 2013]. However, if the signal does not penetrate sufficiently into the canopy, distinguishing between flooded and non-flooded pixels becomes difficult.

For low or sparse vegetation, C-band sensors have been shown to penetrate the canopy with an increase in backscatter as expected [Lang et al., 2008]. Similarly, other studies demonstrated the capability to map floods in short vegetated areas using X-band sensors [Pulvirenti et al., 2011; Pierdicca et al., 2017]. However, the backscatter intensity of flooded vegetation can resemble that of urban areas or moist bare soils, leading to potential misclassifications [Pulvirenti et al., 2015].

Despite these complexities, vegetated areas are usually treated with the same classification techniques as open water surfaces, using backscatter intensity [Cian et al., 2018a; Cazals et al., 2016; Grimaldi et al., 2020]. However, rather than detecting regions of low backscatter, the objective is to identify an increase in backscatter values, which may indicate the presence of flooding. To enhance the classification accuracy of flooded vegetation, new methods employ InSAR to detect changes in the phase of the signal, thereby spotting alterations in surface conditions and improving classification compared to traditional methods [Refice et al., 2014; Pulvirenti et al., 2015]. A review of the classification of vegetated areas is available in Tsyganskaya et al. [2018] and Grimaldi et al. [2020], highlighting both traditional and coherence-based approaches.

2.2.5.3 Urban areas

In urban areas, detecting water is often challenging due to the complex structures of the built environment, including buildings, roads, and various infrastructures. Urban areas typically appear very bright in satellite images due to the double-bounce effect on vertical structures (e.g., buildings, walls) and the high dielectric constants of materials like metal. Smooth surfaces, such as asphalt roads, may exhibit a low backscatter intensity, similar to that of water surfaces. Geometric distortions such as layover and shadows are often present, caused by the height and density of buildings. Additionally, the urban features are often smaller than the sensor resolution, which limits the ability to accurately interpret flood extents in narrow streets or small alleys [Mason et al., 2012b].

Due to these limitations, flood mapping classification in urban areas is more difficult and generally less accurate than in rural or open environments. Traditional classification techniques, which are well-suited for open water detection, underperform in dense urban areas. For instance, Mason et al. [2009b] and Giustarini et al. [2012] applied intensity-based methods, such as region growing, which yielded good results in non-densely built-up areas but resulted in a systematic underdetection of flooded areas in densely urban zones due to geometric effects.

To mitigate this effect, recent research has explored the use of InSAR. As discussed in Section 2.2.3.2 and Equation (2.24), the satellite emits a planar wave with a backscatter intensity and signal phase. Comparing the phase values between two SAR images acquired over the same area at different times allows the calculation of interferometric coherence. In urban areas, high interferometric coherence usually corresponds to stable structures like buildings or roads [Chini et al., 2019]. When comparing a non-flooded and flooded image, it causes low interferometric coherence because of the phase changes. Then, low interferometric coherence can be used as an indicator of flood presence [Giustarini et al., 2012; Pulvirenti et al., 2015; Chini et al., 2019].

2.2.5.4 Post-processing of flood extent

After segmenting flood extent from SAR images, post-processing is often applied to enhance the accuracy of the flood extent. Common post-processing techniques include morphological operations, filtering based on object size, and the integration of ancillary data such as DEM [Martinis et al., 2009].

Morphological operations are used to smooth flood boundaries and fill small holes within the flood extent, yielding a more physically consistent flood extent. They can also be used to remove small and isolated water patches that are unlikely to be part of the flood, based on a minimum flood size ranging from 1,000 m^2 to 10,000 m^2 , depending on the applications.

Moreover, incorporating topographic information from DEM can eliminate falsely detected flooded areas located on elevated terrain where water propagation is unlikely [Pulvirenti et al., 2011]. Other studies refine flood extent by excluding areas not prone to flooding because of excessive slopes or based on the height above the nearest drainage [Twele et al., 2016; Clement et al., 2018; Zhao et al., 2021]. To mitigate over-detection in urban and mountainous regions, areas with shadow and layover regions can be excluded [Huang et al., 2017; Mason et al., 2018]. Land use data are also used to exclude erroneously classified flooded pixels on roads or infrastructures [Tanguy et al., 2017; Grimaldi et al., 2020]. Other recent studies attempt to derive exclusion masks from the SAR images. Martinis et al. [2018] developed a sand exclusion layer to mask out arid regions with low backscatter that are often classified as water. Bauer-Marschallinger et al. [2018] proposed to mask pixels with low backscatter variability in urban areas. These are examples, but multiple exclusion masks exist focusing on specific conditions. Zhao et al. [2021] developed a general exclusion map to automatically identify all pixels that cannot be classified as flooded or non-flooded using SAR backscatter alone (shadow and layover caused by mountains/buildings, sand areas, permanent water bodies, and densely vegetated areas). This exclusion map is available on the Global Flood Monitoring product from the Copernicus programme (<https://global-flood.emergency.copernicus.eu/glofas-forecasting/>).

Note: The role of morphological operations to improve the delineation of flood maps is studied in Chapter 4 to study the uncertainties due to the choice of flood mapping configuration, including post-processing. The role of exclusion layers was not investigated in this work during the calibration of the models, however, it is promising for not taking into account information in specific uncertain areas. The exclusion map from the Global Flood Monitoring product was used in one of the methodologies presented in Chapter 4 to estimate a water depth field from satellite observations.

2.2.5.5 Water depth estimation

Water depths can be estimated from SAR satellite data with InSAR or by combining flood extent with DEM. Another option with less spatial coverage is to resort to altimeters as described in Section 2.2.2.2. A brief overview of the direct retrieval of water depth is provided, as well as its derivation from flood maps.

SAR Interferometry Spatio-temporal changes in water level can be detected using InSAR, exploiting phase differences between two or more acquisitions over the same area:

$$\Delta\Phi = \frac{4\pi}{\lambda}\Delta R, \quad (2.28)$$

where λ is the wavelength and ΔR is the distance change between images to the target in the LOS direction. The water level change is then obtained:

$$\Delta h = \frac{\Delta R}{\cos(\alpha_i)}. \quad (2.29)$$

Over open water surfaces, InSAR is usually ineffective due to specular reflection, resulting in low coherence between the signals. However, in vegetated floodplains or urban areas, the signal returns to the antenna and the signal coherence remains between acquisitions, allowing for interferometric analysis. This method can detect changes in water level with a precision of up to several decimeters or meters. For instance, it was possible to retrieve water levels in the Amazon floodplain up to decimeter precision thanks to InSAR [Aldorf et al., 2001, 2007a].

Derivation from flood maps In Section 2.2.5.1, the flood extent classification of SAR images was introduced. The estimated flood extent can be combined with ancillary data such as DEM to estimate water depths.

In flood extent maps, the easiest feature to identify is the contour corresponding to the interface between wet and dry pixels where $h = 0$. Early methods in the late 20th century overlaid flood maps with topographic contours to determine the water height on the banks of the river [Oberstadler et al., 1997; Brakenridge et al., 1998], yielding a high RMSE of up to 2 m across the flood extent. This procedure is schematically visualized in Figure 2.17.

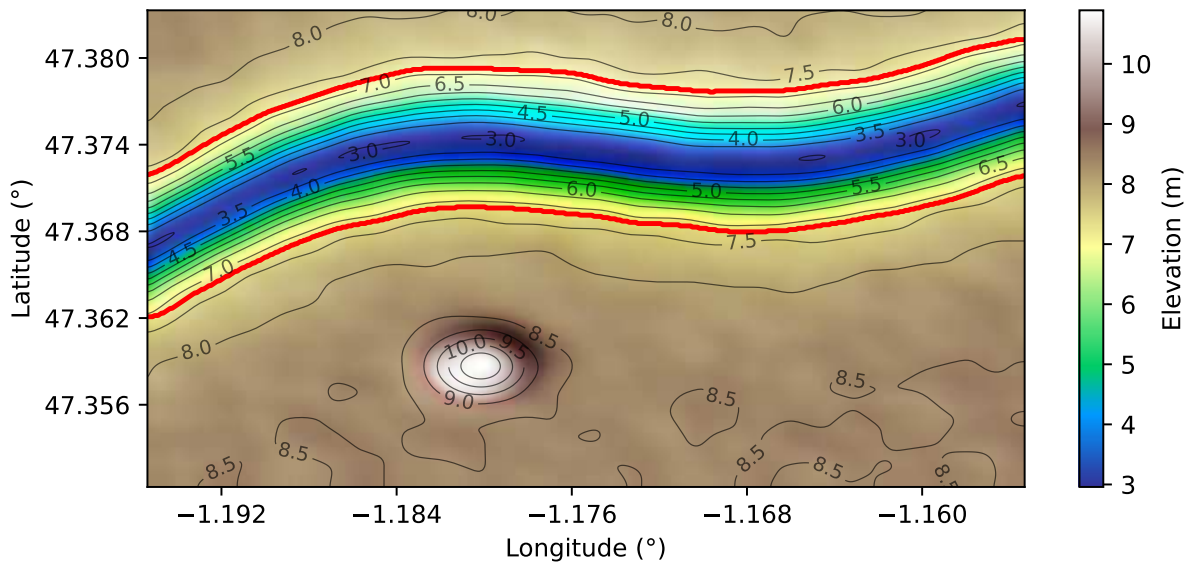


Figure 2.17: Overlay of a flood map (red contour) on a Digital Elevation Model with contour lines every 0.5 m of elevation.

Based on the elevation of the contour where $h = 0$, it is possible to reconstruct the water depth field within the flood extent with simple assumptions such as a flat free surface. Automatic methods, such as the FwDET, quantify water depth over the entire flood extent, considering that the water is, on average, flat. Starting from the border of the flood extent, each pixel within the flood extent is iteratively assigned the elevation of the nearest boundary pixel [Cohen et al., 2018, 2019]. A smoothing step is then performed to reduce artifacts from DEM-flood extent mismatches, though this may also blur real topographic variations. Improvements to this method, incorporating

additional data such as exclusion zones and permanent water masks, were introduced in the FLEXTM methodology, reducing uncertainties by almost half [Betterle and Salamon, 2024]. Several studies then proposed determining the water depth at several cross-sections along a river, given the edges of the flood extent on each cross-section and the assumption that the free surface is flat, as illustrated in Figure 2.18. Between the cross-sections, the water depth is interpolated [Schumann et al., 2007a].

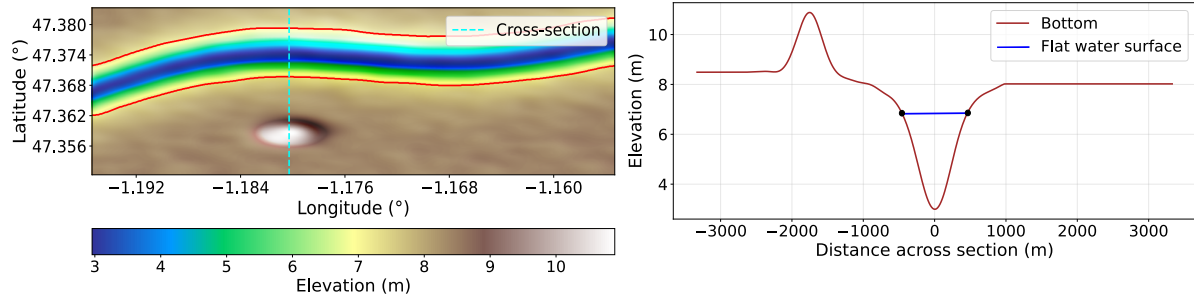


Figure 2.18: Example of schematic cross-section analysis with a Digital Elevation Model and a flood map (red contour) to estimate the water depth.

To introduce physics in the water depth estimation to avoid non-physical estimates, some techniques aimed to introduce hydraulic constraints, treating flood patches as flat and enforcing negative energy gradients from upstream to downstream of the river, though they require detailed field knowledge [Raclot et al., 2003; Raclot, 2006; Puech et al., 2007; Hostache et al., 2009]. Other methods use statistical analysis of boundary elevations [Cian et al., 2018b; Brown et al., 2016] or determine the flood extent by finding the best fit with simulated floods [Nguyen et al., 2016].

The uncertainty in water depth estimation is primarily due to the accuracy of flood extent delineation and the DEM. Water depth uncertainty is assessed using ensemble methods to generate various flood maps using different processing (e.g., speckle filtering, flood extent methods, and hyperparameters, etc.) [Schumann et al., 2008a,c,d; Zwenzner and Voigt, 2009; Schumann et al., 2011]. Table 2.9 summarizes the attempts to derive water depths from SAR satellite data.

Note: The derivation of water depth fields from flood maps is conducted in Chapter 4. In particular, the variability of the water depth estimates across different methods and under various hyperparameters that require tuning is evaluated. The influence of the input flood maps, generated using the methods described in Section 2.2.5.1, on the water depth estimates is evaluated. For generating the water depth fields, SAR images are first preprocessed to reduce speckle noise, after which flood maps are derived and used to produce the water depth fields. All these steps introduce uncertainties on the outputs, which must be quantified to calibrate the 2D flood model.

Table 2.9: Water depth retrieval from flood extents with their respective accuracies and validation data.

Method	Validation data	Error (RMSE)	Reference
Flood contours + DEM	Field / Simulation	0.5–2 m	Oberstadler et al. [1997]; Brakenridge et al. [1998]
Shorelines + simulation	Field	0.2–0.3 m	Mason et al. [2001]
Cross-sections + LiDAR	High water marks	0.2–0.7 m	Matgen et al. [2007a]; Schumann et al. [2007a]
Contours / SRTM / LiDAR + regression	1D Simulation	0.35–1.07 m	Schumann et al. [2008b]
Hydraulic coherence + LiDAR	Field	0.3 m	Hostache et al. [2009]
Flood edges + LiDAR	Aerial data	<0.5 m	Mason et al. [2009a]
Delaunay interpolation + LiDAR	LiDAR	0.15 m	Brown et al. [2016]
Simulated rain fit to extent	Field	<0.2 m	Nguyen et al. [2016]
FwDET + LiDAR	Simulation	<0.7 m	Cohen et al. [2018, 2019]; Peter et al. [2020]
Statistical boundary elevation	Simulation	0.5–0.8 m	Cian et al. [2018b]
DEM + nearest neighbor + HAND	Simulation	<0.5 m	Bryant et al. [2022]

2.3 Summary

Flood events are commonly modeled using 2D SWE. These models should be calibrated to fit the observations and improve their forecast capacity. A classical hypothesis in 2D flood modeling is that the primary source of uncertainty is the bottom shear stress. This term physically represents the resistance to flow due to interactions between the flow and its boundaries. However, in practice, this term accounts for other uncertainties in the model, such as topography, discharge, and boundary conditions.

A common approach in engineering studies is to model the bottom shear stress using the Manning-Strickler formula. The parameter of this formulation (the Strickler or Manning values) is typically spatially variable but assumed to be time-invariant during a flood event. While this friction term should be calibrated, prior information is often available to inform the roughness values. The spatial distribution of friction is based on land cover classification, and each homogeneous subdomain of land use in floodplains is assigned a roughness range using standard tabulated values. It advocates for using UQ to evaluate the impacts of the Strickler parameter on the model outputs as described in Chapter 3 and applied in Chapters 6 and 7.

Satellite observations can provide valuable data for calibrating roughness values. A variety of sensors, including optical, radar, and altimeters, are available. Among these, SAR is the most commonly used for flood monitoring. SAR sensors can monitor floods day and night, even through clouds, with high spatial resolution. Despite the low revisit time, its spatially distributed observations can be used to calibrate roughness in the floodplains.

SAR images are difficult to interpret because of the complexity of signal interactions, which depend on parameters such as incidence angle, surface characteristics, wavelength, geometric effects, or speckle noise. In rural areas, detecting water is generally easier, as it appears as dark regions in SAR images. This is because water reflects the radar signal away from the sensor, producing low backscatter. Several methods are available in the literature to generate flood maps and estimate water depth fields, though with uncertainties. These uncertainties are quantified in Chapter 4 by evaluating the impact of each processing step used to transform raw SAR images into flood maps or water depth estimates. In particular, this chapter highlighted the need to evaluate the impact of speckle reduction, flood mapping, and water depth estimation methods, and the sensitivity of results to hyperparameter tuning.

This chapter highlights, first, that 2D flood models are uncertain because parameters such as roughness are based on expert judgment and implicitly absorb other model errors. Secondly, satellite-derived observations are also uncertain, as their accuracy depends on the processing applied to raw imagery. Therefore, DA provides a framework to calibrate the model parameters through a compromise between the uncertain numerical model and observations. The DA methods used in this thesis are presented in Chapter 3 and applied in Chapters 6 and 7.

Chapter 3

Uncertainty Quantification and Data Assimilation

Résumé :

Deux sources d'information sont disponibles pour améliorer la prévision des inondations futures : les observations (par exemple, dérivées des observations par satellite) et les modèles numériques. Les observations sont sujettes à des erreurs de mesure et de traitement, tandis que les modèles reposent sur des simplifications et des paramètres incertains. Ces deux sources d'information doivent être combinées pour calibrer le modèle numérique, tout en tenant compte de leurs erreurs respectives. Pour relever ce défi, nous présentons dans ce chapitre la quantification d'incertitude et l'assimilation de données qui fournissent des cadres méthodologiques adaptés. Les techniques de quantification d'incertitudes permettent de propager l'incertitude à travers le modèle et d'identifier les paramètres les plus influents par le biais d'une analyse de sensibilité afin de simplifier le problème d'optimisation. L'assimilation de données nous permet de combiner et de trouver un compromis entre les observations et les modèles numériques. La section 3.2 est consacrée aux méthodes de quantification d'incertitude et couvre la propagation d'incertitude et l'analyse de sensibilité. La section 3.3 introduit les principaux concepts du cadre de l'assimilation de données. La section 3.4 présente le problème de l'estimation des paramètres avec une approche bayésienne ainsi que des solutions analytiques (par exemple, meilleur estimateur linéaire sans biais). La section 3.5 présente l'approche bayésienne pour des problèmes dépendant du temps (par exemple, filtre de Kalman, filtre particulaire), et la section 3.6 décrit les approches variationnelles. Les éléments de ce chapitre se sont appuyés sur les revues, cours et thèses suivants [Nodet \[2005\]](#); [Blayo et al. \[2011\]](#); [Robinson \[2015\]](#); [Feyeux \[2016\]](#); [Carrassi et al. \[2018\]](#); [Mouradi \[2021\]](#); [Rouzies \[2023\]](#).

Abstract:

In the context of the thesis, two sources of information are available to improve the forecasting of future flood events: observations (e.g., derived from satellite observations) and numerical models. The observations are subject to measurement and processing errors, while models rely on simplifications and uncertain parameters. These two sources of information should be combined to calibrate the numerical model, while accounting for their respective errors. To address this challenge, we present in this chapter UQ and DA, which provide powerful methodological frameworks. UQ techniques allow to propagate uncertainty through the model, and identify the most influential parameters via sensitivity analysis to simplify the optimization problem. DA allows to combine and find a compromise between observations and numerical models. Section 3.2 is dedicated to UQ, covering uncertainty propagation and sensitivity analysis. Section 3.3 introduces the main concepts of the DA framework. Section 3.4 presents the parameter estimation problem with a Bayesian approach and analytical solutions (e.g., BLUE). Section 3.5 presents Bayesian approaches for time-dependent problems (e.g., KF and PF), and Section 3.6 describes variational approaches. The elements of this chapter are supported by the following reviews, courses, and theses: [Nodet \[2005\]](#); [Blayo et al. \[2011\]](#); [Robinson \[2015\]](#); [Feyeux \[2016\]](#); [Carrassi et al. \[2018\]](#); [Mouradi \[2021\]](#); [Rouzies \[2023\]](#).

Contents

3.1	General notations	64
3.2	Uncertainty Quantification	65
3.2.1	Steps of an Uncertainty Quantification study	65
3.2.2	Uncertainty Propagation	66
3.2.3	Sensitivity Analysis	68
3.3	Introduction to Data Assimilation	74
3.3.1	State of the system	74
3.3.2	True state	75
3.3.3	Model	75
3.3.4	Background knowledge	76
3.3.5	Observations and observation operator	77
3.3.6	Analysis	77
3.4	Bayesian approach	77
3.4.1	Bayesian formulation of DA	77
3.4.2	Best Linear Unbiased Estimator	78
3.5	Bayesian approach for time-dependent models	79
3.5.1	Kalman Filter	79
3.5.2	Extended Kalman Filter	81
3.5.3	Ensemble Kalman Filter	81
3.5.4	Particle Filter	84
3.6	Variational approaches	85
3.6.1	Variational formulation	85
3.6.2	Three-Dimensional Variational formulation	86
3.6.3	Four-Dimensional Variational formulation	87
3.6.4	Incremental Four-Dimensional Variational formulation	88
3.6.5	Error covariance matrices: $\mathbf{B}$ and $\mathbf{R}$	89
3.6.6	Solutions to the variational problems	90
3.7	Summary	92

3.1 General notations

Before introducing UQ and DA frameworks, general notations are introduced and used throughout this chapter. While this chapter presents a general overview of UQ and DA, the notations are illustrated with the quantities they represent in the context of this thesis.

The aim is to calibrate a spatially roughness field represented by a vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{N_p}]$. Each parameter $(\theta_i)_{1 \leq i \leq N_p}$ represents typically a Strickler coefficient corresponding to one of the N_p roughness subdomains defined in the study area based on the land use classification. The flood model propagates the uncertain parameters through time using a model $\mathcal{M}$ to simulate velocities and water depths. In this work, the state variable $\mathbf{x} \in \mathbf{R}^{N_x}$ refers to the simulated water depths, with the following dynamics $\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k, \boldsymbol{\theta}_k)$. To calibrate the uncertain parameters, the state variables are compared to observations $\mathbf{y}^{obs} \in \mathbf{R}^{N_{obs}}$. The observations used in this work are either flood maps (denoted in the following chapters as $O \in \mathbf{R}^{N_x \times N_y}$) consisting of matrices of 0 and 1 or water depth observations in the floodplains extracted from a twin experiment or satellite observations. For comparing $\mathbf{x}$ and $\mathbf{y}^{obs}$, an observation operator $\mathcal{H} : N_x \rightarrow N_{obs}$ transforms the state variables in the observation space (e.g., transformation of the simulated water depth field into a flood map, or interpolation at the water depths locations). More generally, we introduce the composition $\mathcal{G} = \mathcal{H} \circ \mathcal{M}$ to directly map from the parameters to the observations.

In the UQ framework, we consider that Y represents a quantity of interest derived from the model outputs with an operator φ such as $Y = \varphi(\mathcal{G}(\boldsymbol{\theta}))$. For a non-scalar quantity of interest, the vector is denoted as $\mathbf{Y}$. For instance, in Chapters 5 the quantity Y of interest is the evaluation of a performance measure between the model outputs $\mathcal{G}(\boldsymbol{\theta})$ and observations $\mathbf{y}^{obs}$, and in Chapter 6, the quantities of interest are directly the results of $\mathcal{G}(\boldsymbol{\theta})$, the water depth at the observation locations.

For the DA framework, we introduce additional notations for the parameters and state variables with superscripts b , a , f , and t corresponding to background, analysis, forecast, and true state, respectively. The background $\boldsymbol{\varepsilon}^b$ and observation errors $\boldsymbol{\varepsilon}^o$ are modeled with Gaussian variables with error covariance $\mathbf{B}$ and $\mathbf{R}$, respectively. An additional under-script k is added for time-dependent problems, with, for instance, $\mathbf{x}_k$ representing the state variable at time t_k .

The principal notations used in this chapter are summarized in Table 3.1.

Table 3.1: Summary of symbols used in UQ/DA and their context in the thesis.

Symbol	Meaning in UQ/DA	Description/Context in thesis
$\boldsymbol{\theta} \in \mathbf{R}^{N_p}$	Uncertain parameters	Roughness coefficients
$\mathbf{x} \in \mathbf{R}^{N_x}$	State variables	Modeled water depths
$\mathbf{y}^{obs} \in \mathbf{R}^{N_{obs}}$	Observations	Flood maps or water depths
$\mathcal{M}$ or $\mathbf{M}$	Numerical model	2D shallow-water solver
(linearized)	$\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k, \boldsymbol{\theta}_k)$	(TELEMAC-2D)
$\mathcal{H}$ or $\mathbf{H}$	Observation operator	Maps state variables to observation space
(linearized)		
$\mathcal{G} = \mathcal{H} \circ \mathcal{M}$ or $\mathbf{G}$ (linearized)	Composite operator	Maps parameters to observation space
$Y = \varphi(\mathcal{G}(\boldsymbol{\theta}))$	Quantity of interest (scalar)	Water depth at single points/ performance measures evaluation
$\mathbf{Y}$	Quantities of interest (vector)	Water depth at several location or times

Symbol	Meaning in UQ/DA	Description/Context in thesis
$\mathbf{x}^b, \boldsymbol{\theta}^b$	Background state/parameter	Prior guess of roughness values before assimilation
$\mathbf{x}^a, \boldsymbol{\theta}^a$	Analysis state/parameter	Updated estimate after assimilation
$\mathbf{x}^f$	Forecast state	Model propagation before update
$\mathbf{x}^t, \boldsymbol{\theta}^t$	True state	Unknown true water depth field/roughness field
$\boldsymbol{\varepsilon}^{\text{obs}} \sim \mathcal{N}(0, \mathbf{R})$	Observation error	SAR classification uncertainty
$\boldsymbol{\varepsilon}^b \sim \mathcal{N}(0, \mathbf{B})$	Background error	Prior uncertainty on roughness
$\mathbf{B}$	Background error covariance	Uncertainty in background roughness
$\mathbf{R}$	Observation error covariance	Uncertainty in water depth or flood maps
$\mathbf{Q}$	Model error covariance	Neglected

UQ and DA are presented within a probabilistic framework. For the uncertain parameters, state variables, and observations, the probability density functions are denoted as $p(\boldsymbol{\theta})$, $p(\mathbf{x})$, and $p(\mathbf{y}^{\text{obs}})$, respectively. To characterize the probability density functions, statistical moments can be used. The k -th moment of a random variable Z is defined as:

$$m_k = \mathbb{E}[Z^k] = \int Z^k p(Z) dZ. \quad (3.1)$$

The first moment is the expectation of Z , denoted $\mathbb{E}[Z]$. Its empirical estimate over a sample $\{Z^{(i)}\}_{i=1}^N$ of size N is the statistical average denoted as $\bar{Z}$ and calculated as:

$$\mathbb{E}[Z] = \bar{Z} = \frac{1}{N} \sum_{i=1}^N Z^{(i)}. \quad (3.2)$$

Similarly, the second central moment is the variance of Z denoted $\text{Var}(Z) = \mathbb{E}[(Z - \mathbb{E}[Z])^2] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$. Its empirical estimate is written as:

$$\text{Var}[Z] = \frac{1}{N-1} \sum_{i=1}^N (Z^{(i)} - \bar{Z})^2. \quad (3.3)$$

The standard deviation is derived from the variance as $\sigma_Z = \sqrt{\text{Var}(Z)}$. The covariance of two components Z_1 and Z_2 is the operator defined as:

$$\text{Cov}(Z_1, Z_2) = \mathbb{E}[(Z_1 - \mathbb{E}[Z_1])(Z_2 - \mathbb{E}[Z_2])]. \quad (3.4)$$

Covariance can be empirically estimated as:

$$\text{Cov}(Z_1, Z_2) = \frac{1}{N-1} \sum_{i=1}^N (Z_1^{(i)} - \bar{Z}_1)(Z_2^{(i)} - \bar{Z}_2). \quad (3.5)$$

3.2 Uncertainty Quantification

3.2.1 Steps of an Uncertainty Quantification study

Numerical models aim to simulate real phenomena (e.g., floods) by solving equations that describe the dynamics of a system. However, they rely on simplifying assumptions (physical and numerical approximations), numerical schemes, and uncertain input parameters,

so they never perfectly reproduce real phenomena. The objective of UQ studies is to quantify the impact of these modeling uncertainties (often focusing on the uncertainties in the input parameters) on a simulated quantity of interest [de Rocquigny et al., 2008; Pasanisi and Dutfoy, 2011].

Figure 3.1 reports the general workflow of UQ studies consisting of:

- Step A: Definition of the uncertain input parameters $\boldsymbol{\theta} \in \mathbb{R}^{N_p}$, the numerical model $\mathcal{M}$, the outputs $\mathcal{G}(\boldsymbol{\theta}) = \mathcal{H}(\mathcal{M}(\boldsymbol{\theta}))$ with $\mathcal{H}$ an observation operator transforming the model outputs into the observation space, and a quantity of interest $Y = \varphi(\mathcal{G}(\boldsymbol{\theta}))$ extracted from the outputs with an operator φ . The distribution of the quantity of interest is derived from Y either with statistical analysis (mean, variance, quantiles, etc.) or by fitting empirically a probability distribution with kernel methods [Silverman, 2018].
- Step B: The probability density functions of the input parameters $\boldsymbol{\theta}$ are quantified thanks to scientific literature, statistical fitting with in-situ measurements, or expertise. Usually, the parameters are considered to be independent, but sometimes correlations between the parameters should be specified and taken into account.
- Step C: Uncertainty Propagation consists in applying the numerical model with uncertain input parameters via various methods (e.g., Monte Carlo sampling strategies) to estimate the distribution of the quantity of interest.
- Step D: SA quantifies the input's influence on the quantity of interest variability. The less influential parameters are then fixed, and the most influential ones are calibrated. It reduces the complexity of calibration or DA problems by decreasing their dimensionality.
- Step E: The input parameters $\boldsymbol{\theta}$ are estimated using model outputs and observations. This step is conducted using DA techniques in this work, but other parameter estimation techniques can be used.

This section focuses on Step C and Step D of the UQ methodology. Section 3.2.2 describes the uncertainty propagation step. Then, Section 3.2.3 presents several algorithms for sensitivity analysis.

Note: Chapter 2 highlighted the need to conduct a UQ study with the roughness field to estimate the impact of its uncertainties in flood models on the outputs. Furthermore, we noted that the main outputs of satellite observations are water depth fields or flood maps. Then, in this work, the quantity of interest will be either water depths at single or several points and the computation of the similarity between the flood maps and a reference (through what we will call a performance measure in Chapter 5). The other sources of uncertainty in the model (e.g., boundary conditions, discretization, modeling simplifications, topography) are not considered, as we assume that the roughness field should compensate for these errors. Furthermore, the inlet discharge and the topography are considered to be perfectly known and noise-free.

3.2.2 Uncertainty Propagation

In a typical Monte Carlo approach, Uncertainty Propagation aims to propagate the uncertain input parameters $\boldsymbol{\theta}$ with the numerical model and quantify the quantity of interest Y . The uncertain input parameters are defined with prior probability distributions, and

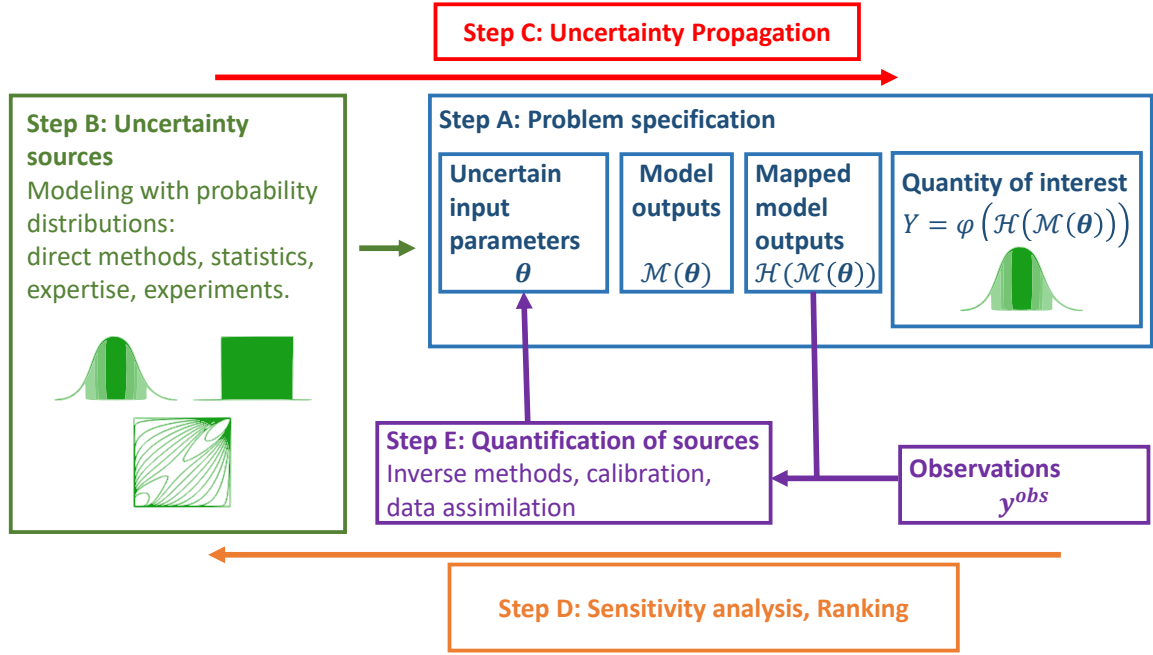


Figure 3.1: Sketch representing the different steps in an Uncertainty Quantification study adapted from Baudin et al. [2015].

N combinations of input parameters are generated (called DoE) to evaluate the model at these points. The number of generated combinations should be high enough to ensure convergence of the N -sample towards the distribution of the quantity of interest, but not too high to prevent from computational expensive costs. There are various approaches for generating DoE, such as:

- Random sampling: The most well-known and simplest technique is random sampling, where samples are generated randomly from their probability distributions using the Monte-Carlo method [Metropolis and Ulam, 1949]. Random sampling may leave gaps in the parameter space, especially when the number of parameters increases. To mitigate this problem, alternatives are possible, such as stratifying techniques.
- LHS: In LHS, each input parameter range is divided into N equally probable intervals to cover the whole parameter space [McKay et al., 2000]. Then, one value is randomly selected from each interval without replacement. For a N_p -dimensional input space, there are $(N!)^{N_p}$ possibilities to build the experimental design. An extension of LHS is nested LHS [Meliani et al., 2019], which allows for successive refinements of a sample set while preserving stratification properties. It ensures that new samples added to an existing LHS design maintain the space-filling characteristics across multiple levels of fidelity.
- Low-discrepancy sequences: Let $\{\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}\}$ be a sequence of N points in the unit cube $I_{N_p} = [0, 1]^{N_p}$ and J a sub-interval of I_{N_p} . Then, the discrepancy is given by:

$$D^*(\{\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}\}) = \sup_J \left| \frac{A(J)}{N} - V(J) \right|, \quad (3.6)$$

where D^* measures the maximum deviation between the N points and the uniform distribution over sub-rectangle of the unit hypercube, $A(J)$ is the number of points

$\theta^{(i)}$ in J , and $V(J)$ the volume of J . Low discrepancy sequences are constructed to minimize the discrepancy D^* [L'Ecuyer and Lemieux, 2002].

Figure 3.2 reports 3 DoE with random sampling, LHS, and low-discrepancy sequences generated with the OpenTURNS (an Open source initiative for the Treatment of Uncertainties, Risks'N Statistics) Python library (<https://openturns.github.io/www/>). For random sampling, the points are randomly distributed in the search space, and some zones are not explored, while the spread in the search space is better for the other two methods.

The Uncertainty Propagation was described with a Monte-Carlo classical approach. However, alternative uncertainty propagation techniques exist, such as PCE, which approximate the model response as an expansion in orthogonal polynomials of the random inputs. This approach can be more computationally friendly.

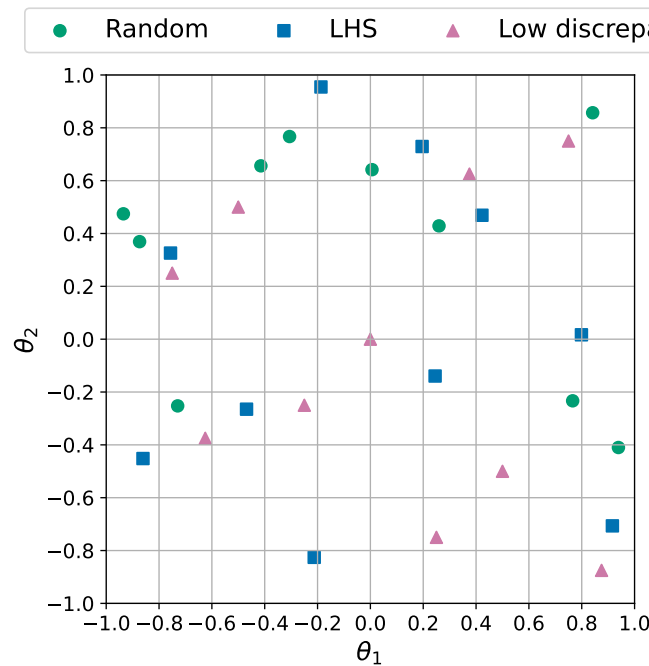


Figure 3.2: Example of Design of Experiment with Random Sampling, Latin Hypercube Sampling, and low discrepancy sequence in the search space $[-1, 1]^2$ with uniform input distributions.

Note: In the present thesis, the DoE is generated with LHS or nested LHS to ensure stratification of the search space. This choice ensures good coverage of the high-dimensional uncertain parameters (more than 90 in one of the study areas). It limits the number of required simulations, given that the flood model is costly in this case.

3.2.3 Sensitivity Analysis

SA is used for various applications for modeling physical systems [Saltelli et al., 2004]. SA improves the understanding of the numerical model by studying the relationship between the input and output, or by defining critical zones in the parameter space that lead to extreme outcomes. It is often used to simplify the model by fixing non-influential parameters to reduce the dimensionality of the input space or to construct a metamodel. SA methods dichotomized into two approaches: LSA and GSA.

- LSA analyzes the variation of the quantity of interest for a slight variation of the input parameter around a reference value. The analysis is computed with partial derivatives of the models around the reference value with adjoint differentiation methods or finite-difference approximation methods [Cacuci, 2003]. This enables analyzing the behavior of the numerical model in the neighborhood of the reference value, but may not apply to the entire input space, especially when the numerical model is nonlinear.
- GSA, on the other hand, explores the variations of the quantity of interest for the whole input space [Saltelli et al., 2008]. GSA can be performed to identify influential from non-influential parameters (screening) or to rank the input parameters depending on their influence on the quantity of interest. There exists a wide variety of GSA methods, and the choice of method highly depends on the numerical model properties (e.g., number of parameters, dependence of the input parameters, linearity of the numerical model, computation cost). For interested readers, reviews of the different sensitivity analysis methods used in the context of environmental models are available in Saltelli et al. [2000], Iooss and Lemaître [2015], and Pianosi et al. [2016].

Screening techniques (e.g., Morris method), variance-based decomposition (e.g., Sobol's indices), and moment-independent measures are commonly used in GSA. These methods are presented in the following sections. Most of the methods were developed for a scalar quantity of interest. Then, the quantity of interest Y is a scalar output (e.g., water depth at a point and for a given time, or performance measure evaluation between two flood maps).

Note: In the present thesis, GSA is preferred over LSA since the flood model is nonlinear and contains a larger number of uncertain parameters.

3.2.3.1 Screening methods

Screening methods determine in a few model evaluations, the behavior of the model and identify the non-influential parameters [Da Veiga et al., 2021]. Morris screening method [Morris, 1991; Faivre et al., 2013] consists in computing elementary effects, i.e., quantifying output variations between successive input points (on trajectories). The elementary effects are averaged over multiple trajectories to provide a more global and robust assessment of input sensitivities.

The input parameter space $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_{N_p}\}$ is mapped to a hypercube $[0, 1]^{N_p}$, and discretized into a grid with L levels with a discretization step Δ . The intersection of the different levels defines an ensemble of L^{N_p} nodes. Morris screening method consists in moving each parameter by a step Δ one at a time along a trajectory t starting from a random node of the discretized hypercube. Then, the elementary effect of an input parameter θ_i for trajectory t writes as:

$$EE_i^{(t)} = \frac{Y(\theta_1, \theta_2, \dots, \theta_i + \Delta, \dots, \theta_{N_p}) - Y(\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_{N_p})}{\Delta} \quad (3.7)$$

Elementary effects are computed along r trajectories. Examples of trajectories in two dimensions with $L=9$ levels are reported in Figure 3.3.

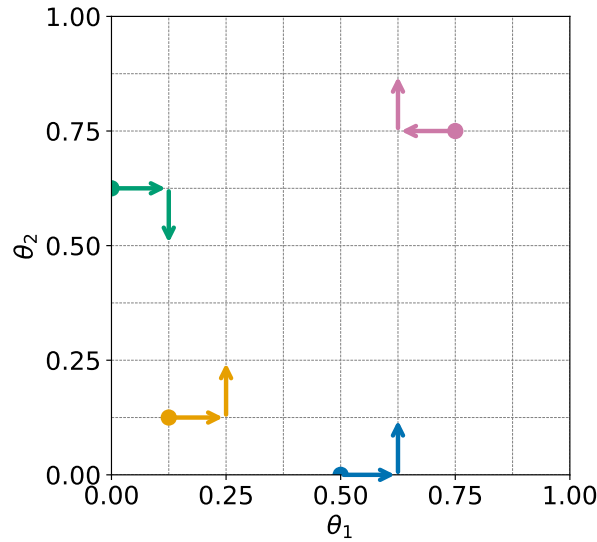


Figure 3.3: Example of $r = 4$ trajectories in two dimensions on the Hypercube $[0, 1]^2$ with $L=9$ levels and a discretization step of $1/8$. The filled circles represent the start of the trajectories.

A statistical interpretation of the elementary effects can be computed with the mean and variance for each input parameter on the r trajectories, such as:

$$\mu_i^* = \frac{1}{r} \sum_{t=1}^r |EE_i^{(t)}| \quad \text{and} \quad (\sigma_i^*)^2 = \frac{1}{r-1} \sum_{t=1}^r \left(EE_i^{(t)} - \mu_i^* \right)^2 \quad (3.8)$$

μ^* is considered low when it is close to zero (e.g., less than 5 to 10% of the variation range of Y), and high when approaching the total variation range of Y . Similarly, σ^* is considered low when it is much smaller than μ^* and high when it is of the same order of magnitude as μ^* . Then, three types of input parameters can be identified:

1. Low μ^* , and low σ^* : Inputs with negligible influence.
2. Low μ^* , and high σ^* : Inputs with linear influence without interaction.
3. High μ^* , and high σ^* : Inputs with nonlinear influence and/or with interactions.

Remark: To ensure an unbiased exploration of the input space, the distribution of the sampled points should follow a uniform distribution. Then, it is advised to use a discretization step $\Delta = \frac{L}{2(L-1)}$ and consider at least 10 trajectories [Morris, 1991; Campolongo et al., 2007].

3.2.3.2 Ranking methods for GSA

Screening methods classify parameter inputs as influential or not, but do not rank their influence on the quantity of interest Y . Then, sensitivity methods that quantify the effect of a parameter on the distribution of the quantity of interest must be used, with sensitivity indices [Saltelli et al., 2008]. We first describe regression-based methods, the variance-based Sobol method [Saltelli et al., 2004], and finally introduce the Borgonovo method [Borgonovo, 2007].

Regression-based methods A simple approach for measuring the influence of each input on the output distribution is to consider linear correlations by evaluating the Pearson (r) or the Standardized Regression Coefficient (SRC) defined as:

$$r(\theta_i, Y) = \frac{\text{Cov}(\theta_i, Y)}{\sqrt{\text{Var}[\theta_i]\text{Var}[Y]}} \quad SRC(\theta_i, Y) = \beta_i \sqrt{\frac{\text{Var}[\theta_i]}{\text{Var}[Y]}}, \quad (3.9)$$

where β_i is the estimated regression coefficient for input θ_i .

These coefficients are useful to quantify the influence of the input parameters if the model is linear. For nonlinear but monotone models, rank transformation can be used, leading to SRRC and PRCC. However, for nonlinear and nonmonotonic models, these approaches are often not adapted [Saltelli and Sobol, 1995].

Analysis of variances and Sobol's indices for scalar outputs By decomposing the output variance, these methods quantify the sensitivity of the quantity of interest to each input and their interactions [Saltelli et al., 2004]. The most popular variance-based method is the Sobol method [Sobol, 1993]. This method is based on the ANOVA decomposition of Y . Its decomposition as a sum of elementary functions writes as [Hoeffding, 1992; Sobol, 1993]:

$$Y = f_0 + \sum_{i=1}^{N_p} f_i(\theta_i) + \sum_{i<j}^{N_p} f_{i,j}(\theta_i, \theta_j) + \dots + f_{1,\dots,N_p}(\theta_1, \dots, \theta_{N_p}), \quad (3.10)$$

where f_0 is a constant and the other functions verify:

$$\forall k = 1, \dots, s \forall \{i_1, \dots, i_s\} \subseteq \{1, \dots, N_p\} \int_0^1 f_{i_1, \dots, i_s}(\theta_{i_1}, \dots, \theta_{i_s}) d\theta_{i_k} = 0. \quad (3.11)$$

If Y is square integrable on its domain and if the inputs are independent, then:

$$\text{Var}[Y] = \sum_{i=1}^{N_p} V_i + \sum_{1<j}^{N_p} V_{i,j} + \sum_{i<j<k}^{N_p} V_{i,j,k} + \dots + V_{1,2,\dots,N_p} \quad (3.12)$$

where $V_i(Y) = \text{Var}[\mathbb{E}[Y|\theta_i]]$, $V_{i,j}(Y) = \text{Var}[\mathbb{E}[Y|\theta_i, \theta_j]] - V_i(Y) - V_j(Y)$ etc. Then, Sobol sensitivity indices are naturally defined as:

$$SI_{i_1, \dots, i_s} = \frac{V_{i_1, \dots, i_s}}{\text{Var}[Y]}. \quad (3.13)$$

The first-order Sobol index $SI_i = \frac{V_i}{\text{Var}[Y]}$ only accounts for the variability of the quantity of interest distribution due to the input parameter θ_i without interactions. It gives the percentage of the variance (varying between 0 and 1) that can be reduced if the uncertainty on θ_i is eliminated.

To estimate all the effects of an input on the quantity of interest, the total sensitivity analysis index was introduced [Homma and Saltelli, 1996]:

$$TSI_i = SI_i + \sum_{i \neq j} SI_{i,j} + \sum_{j \neq i, k \neq i, j < k} SI_{i,j,k} + \dots = \sum_{l \in \#i} SI_l, \quad (3.14)$$

where $\#i$ represents all the subsets of indices containing i . The total order Sobol index represents both the direct effect on an input parameter θ_i and its interactions with the other parameters on the output.

The Sobol indices are usually computed with Monte-Carlo methods [Sobol, 1993]. However, the convergence rate is in $\sqrt{N}$, requiring thousands of evaluations of the model.

Sobol indices with PCE For costly numerical models, variance-based methods leverage the construction of metamodels with PCE. Any second-order random variable can be decomposed with PCE in an unique decomposition [Ghanem and Spanos, 2003] such as:

$$Y = \sum_{\alpha \in \mathbb{N}^{N_p}} y_{\alpha} \psi_{\alpha}(\boldsymbol{\theta}) \quad (3.15)$$

where ψ_{α} is a polynomial orthonormal basis and y_{α} are the coordinates of Y in the basis. The PCE can be decomposed similarly to the ANOVA decomposition, and the Sobol' indices are computed based on the post-treatment of the PCE [Sudret, 2008].

The computation of the PCE can be costly as the number of terms grows with the size of the input parameter, so truncation schemes are used to build a sparse PCE. For instance, the LARS method [Efron et al., 2004] was used to automatically detect the PCE's significant coefficients [Blatman and Sudret, 2011].

Aggregated Sobol indices The Sobol indices were designed for scalar outputs. However, the quantity of interest can be a vector, such as water depth observations at multiple points or times. For $\mathbf{Y} = [Y_1, \dots, Y_{N_{obs}}] \in \mathbb{R}^{N_{obs}}$ with N_{obs} the number of observations, Gamboa et al. [2014] proposed an extension of Sobol indices called aggregated Sobol indices. For a subset of input parameters $A = \{i_1, \dots, i_s\}$, they are defined as:

$$ASI_A = \frac{\sum_{j=1}^{N_{obs}} \text{Var}[Y_j](SI_A)_j}{\sum_{j=1}^{N_{obs}} \text{Var}[Y_j]} \quad (3.16)$$

where $\text{Var}[Y_j]$ is the variance of the j -th component of $\mathbf{Y}$ and $(SI_A)_j$ is the Sobol index of Y_j for the subset of input parameters A .

Sobol indices are computed for each component of the output, so for high-dimensional outputs, aggregated Sobol indices are costly. Lamboni et al. [2011] proposed to apply PCA decomposition on $\mathbf{Y}$ to approximate the generalized indices of Gamboa et al. [2014]. Over the DoE with N realizations, the full set of outputs is defined as:

$$\mathcal{Y} = \begin{bmatrix} Y_1^{(1)} & \dots & Y_t^{(1)} & \dots & Y_{N_{obs}}^{(1)} \\ \vdots & & \vdots & & \vdots \\ Y_1^{(i)} & \dots & Y_t^{(i)} & \dots & Y_{N_{obs}}^{(i)} \\ \vdots & & \vdots & & \vdots \\ Y_1^{(N)} & \dots & Y_t^{(N)} & \dots & Y_{N_{obs}}^{(N)} \end{bmatrix}. \quad (3.17)$$

The columns of $\mathcal{Y}$ are normalized and centered such that:

$$(\mathcal{Y}_c)_{t,i} = \frac{Y_t^{(i)} - \mathbb{E}[Y_t]}{\sqrt{\text{Var}[Y_t]}}. \quad (3.18)$$

The correlation matrix of the columns of $\mathcal{Y}_c$ can be decomposed with PCA [Jolliffe and Cadima, 2016] as:

$$\Sigma = \frac{1}{N} \mathcal{Y}_c^T \mathcal{Y}_c = \sum_{k=1}^{N_{obs}} \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \quad (3.19)$$

where $\mathbf{v}_k$ are normalized ($\|\mathbf{v}_k\| = \sqrt{\mathbf{v}_k^T \mathbf{v}_k} = 1$) and mutually orthogonal eigenvectors with eigenvalues λ_k . Then, we have:

$$Y_t^{(i)} = \mathbb{E}[Y_t] + \sum_{k=1}^{N_{obs}} \mathbf{h}_k^{(i)} (\mathbf{v}_k)_t \quad (3.20)$$

where $\mathbf{h}_k = \mathcal{Y}_c \mathbf{v}_k$. Usually, Equation (3.20) is truncated with the J first components of the PCA such that $J \ll N_{obs}$. Finally, the generalized indices for any group of input parameters A are given by:

$$GSI_A = \frac{\sum_{j=1}^J \lambda_j (SI_A)_j}{\sum_{j=1}^J \lambda_j}, \quad (3.21)$$

where $(SI_A)_j$ is the Sobol index of $\mathbf{h}_j$ for the group of input parameters A .

Density-based GSA The Sobol method assumes that the first two moments describe entirely the quantity of interest distribution. However, in some applications, variance is not sufficient. To address this limitation, alternative sensitivity analysis methods were proposed. For instance, the HSIC method computes the covariance between any nonlinear transformation of an input parameter and any nonlinear transformation of the quantity of interest [Gretton et al., 2005]. The Borgonovo method [Borgonovo, 2007] (also called δ -sensitivity measures) quantifies the impact of input parameters on the entire output distribution by comparing the unconditional output probability density function with its conditional counterpart.

Borgonovo method Borgonovo method finds groups of inputs that, if fixed, lead to the most significant modification of the quantity of interest distribution. Let us consider $p(Y)$ the density of Y , and $p(Y|\theta_i)$ the conditional density of Y given that θ_i is fixed. Then, for any subset $A = [i_1, \dots, i_s]$ of input parameters, the L^1 distance between the unconditional output and its conditional version (total area between the two curves) writes as:

$$s(\theta_A) = \int_{\mathbb{R}} |p(Y) - p(Y|\theta_A)| dY. \quad (3.22)$$

Borgonovo's, δ -sensitivity is defined for the group of input parameters θ_A as the expected shifts between the distributions:

$$\delta_A = \frac{1}{2} \mathbb{E}[s(\theta_A)] = \frac{1}{2} \mathbb{E} \left[\int_{\mathbb{R}} |p(Y) - p(Y|\theta_A)| dY \right] \quad (3.23)$$

Compared to Sobol' indices, Borgonovo quantifies any change in the shape of the output distribution. They can be computed from any arbitrary Monte Carlo samples, but can be computationally expensive for high-dimensional systems as they require reliable density estimation. The Sobol and Borgonovo indices values cannot be compared directly, but the ranking of the inputs for both methods reveals whether variance-based methods miss some effects.

Note: In this work, the Sobol indices are used in conjunction with the Borgonovo indices to verify that for two SA methods, the rankings match and ensure that influential parameters are identified as recommended in [Borgonovo, 2007; Pianosi et al., 2016]. Screening methods were not used, as we aim to rank the influential parameters to gain a better understanding of the role of each floodplain subdomain. The methods are applied in Chapters 6 and 7 to determine which floodplain subdomains are influential with different quantities of interest (performance measure evaluation between simulated and observed flood maps for Chapter 6, and simulated water depth fields for Chapter 7).

3.3 Introduction to Data Assimilation

DA can be traced back to the first Numerical Weather Prediction in the early 20th century by Lewis Fry Richardson [Richardson, 1922] to update the state of the atmosphere. In this first thought experiment, Richardson imagined solving the governing equations of the atmosphere step-by-step by combining them with observations. He attempted to make a 6-hour forecast at a point near Munich in Germany by solving an atmospheric model by hand, initialized with a set of available observations. However, his forecast was incorrect, predicting an unrealistic pressure change of over 145 hPa. It was considered the first historical attempt to combine observations with a numerical model. The Second World War accelerated the deployment of global measurement networks, and with the development of the first computers, DA methods were developed in the 1950s and 1960s. While DA was widely used for Numerical Weather Prediction in the 20th century, it is now applied to various geoscience problems [Carrassi et al., 2018] such as flood forecasting [Grimaldi et al., 2016; Schumann et al., 2022], hydrological modeling [Le Dimet et al., 2009], fire propagation [Rochoux et al., 2014; Cheng et al., 2022], or oceanography [Martin et al., 2015; Edwards et al., 2015].

DA can also be thought of as a particular case of an inverse problem that combines observations and a dynamic model with a physical background and includes their respective errors (e.g., the error of representation, measurement error, model errors, etc.) [Asch et al., 2016; Carrassi et al., 2018]. DA can be used to reach different goals:

- Forecasting problem: The future state of the system is predicted based on past and current observations. The forecasting capacity of the system is enhanced by updating it with an optimal initial condition.
- Smoothing or reanalysis problem: The past and current states of the system are updated thanks to past and current observations. It helps analyze the system afterwards and better understand its evolution. This technique is, for instance, used for atmospheric reanalysis of the global climate, such as the ERA5 reanalysis product [Hersbach et al., 2020].
- Filtering problem: Correct the current state of a system from current and past observations.
- Calibration problem: Calibration or estimation problems aim to estimate the true value of the parameters instead of the state variables.

In DA, different types of information coexist and interact. Before introducing the methods to solve the introduced problems, we present the main concepts and notations of DA in Figure 3.4 and in the following subsections:

3.3.1 State of the system

DA aims to estimate the state of a system, which is generally represented in a discrete vector denoted as $\mathbf{x} \in \mathbb{R}^{N_x}$ (e.g., water depth field in this work). The state vector contains physical variables corresponding to one or several quantities, often spatially varying (e.g., water depth, flow velocities, temperature), that the model evaluates using partial differential equations with respect to time. In addition to the state variables, most models depend on parameters that describe the inherent physical properties (e.g., roughness, soil moisture) or empirical coefficients. These parameters, denoted as $\boldsymbol{\theta} \in \mathbb{R}^{N_p}$, are usually uncertain and must be calibrated. For discrete-time dynamics, the state and parameter vectors at time t_k are denoted as $\mathbf{x}_k$ and $\boldsymbol{\theta}_k$.

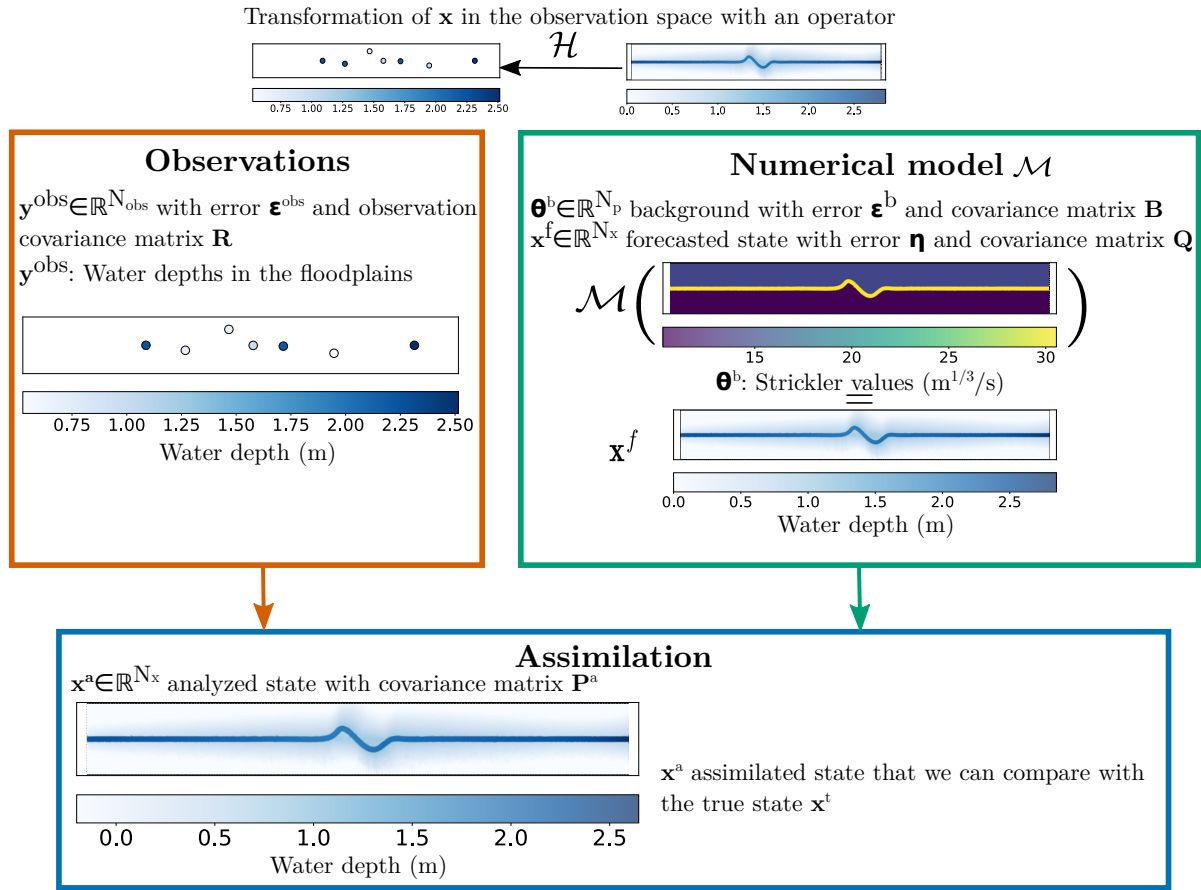


Figure 3.4: Principle of Data Assimilation and notations of the problem on a schematic example of a flood event.

Remark: The state vector $\mathbf{x}$ is linked with the parameters $\boldsymbol{\theta}$ by the model such that $\mathbf{x} = \mathcal{M}(\boldsymbol{\theta})$. The random vector $\boldsymbol{\theta}$ is commonly referred to as the control vector, as it controls the state variables through the numerical model. For 2D flood modeling, the state vector is usually the water depth, the velocities at the mesh nodes or discharges, and the parameters are the roughness values in different subdomains.

For joint state-parameter estimation, the state vector is usually replaced by an augmented state combining both the state variables and the parameters such that:

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{pmatrix} \in \mathbb{R}^{N_x + N_p}. \quad (3.24)$$

3.3.2 True state

The true state of a physical system represents the actual, but generally unknown, values of the state vector. The assimilation process aims to estimate the true state or parameter, denoted as $\mathbf{x}^t$ or $\boldsymbol{\theta}^t$. For evaluating DA algorithms, a usual practice is to use a twin experiment where the true values of the state and parameter vectors are known.

3.3.3 Model

The governing equations of a physical system are solved with a model $\mathcal{M}$ (e.g., TELEMAC-2D flood model). For linear systems, the model is denoted as $\mathbf{M}$. In geoscience, it is

typically a mathematical model solving differential equations, such as:

$$\begin{cases} \frac{\partial \mathbf{x}}{\partial t} = \mathcal{M}(t, \mathbf{x}, \boldsymbol{\theta}) \\ \mathbf{x}(t=0) = \mathbf{x}_0 \end{cases}, \quad (3.25)$$

where $\mathcal{M}$ is a function of time and space, and $\mathbf{x}_0$ is the initial state. In practice, numerical models are usually discrete in time. The model propagates the state from t_k to t_{k+1} according to:

$$\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k, \boldsymbol{\theta}_k). \quad (3.26)$$

However, Equation (3.26) is an idealized representation of the system dynamics. Models are simplifications of the true system, and the integration of the model between t_k and t_{k+1} introduces model errors. A more realistic representation is given:

$$\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k^t, \boldsymbol{\theta}_k^t) + \boldsymbol{\eta}_k, \quad (3.27)$$

where $\boldsymbol{\eta}_k$ is the model error at time t_k . This error originates from multiple sources, including numerical discretization, approximations in representing physical processes, or uncertainties in boundary conditions. $\boldsymbol{\eta}_k$ is often modeled as a Gaussian random variable with mean and model covariance errors:

$$\bar{\boldsymbol{\eta}}_k = \mathbb{E}[\boldsymbol{\eta}_k] \quad \mathbf{Q}_k = \mathbb{E}[(\boldsymbol{\eta}_k - \bar{\boldsymbol{\eta}}_k)(\boldsymbol{\eta}_k - \bar{\boldsymbol{\eta}}_k)^T]. \quad (3.28)$$

In practice, the quantification of model errors is difficult. A common assumption is to neglect model error by setting $\boldsymbol{\eta} = \mathbf{0}$. Furthermore, in parameter estimation problems, there is no time dynamics on the parameters, and a persistence model is considered, such as $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$.

Remark: In the following, the algorithms can be reformulated by replacing $\mathbf{x}$ with $\boldsymbol{\theta}$ and inversely, since we assume a deterministic dynamics $\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k, \boldsymbol{\theta})$, with persistent parameters and no model error. Correcting the parameter is therefore equivalent to correcting the states.

3.3.4 Background knowledge

The background knowledge of the state or parameter is denoted by $\mathbf{x}^b$ (for states) or $\boldsymbol{\theta}^b$ (for parameters). This term integrates all the knowledge before assimilating information. For instance, the information is historical measurements, expert opinions, or the output of the model from the propagation step of the assimilation of the previous window. The background error is modeled as:

$$\boldsymbol{\epsilon}_k^b = \boldsymbol{\theta}_k^b - \boldsymbol{\theta}_k^t. \quad (3.29)$$

The background error is often modeled as a Gaussian random variable:

$$\boldsymbol{\epsilon}_k^b \sim \mathcal{N}(0, \mathbf{B}_k), \quad (3.30)$$

where $\mathbf{B}_k$ is referred to as the background error covariance matrix at time t_k . Background knowledge is crucial in DA, as it acts as an initial guess and as a Tikhonov regularization term to have a well-posed problem [Tikhonov and Arsenin, 1977].

3.3.5 Observations and observation operator

The observation vector of a physical system is denoted as $\mathbf{y}^{obs} \in \mathbb{R}^{N_{obs}}$. Observations originate from various sources and may not correspond to the state variables; they may measure other physical quantities or be located at positions that do not coincide with the discretized model. The observations at time t_k are related to the true system state through an observation operator $\mathcal{H}_k : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_{obs}}$ (for linear models, the observation operator is denoted as $\mathbf{H}$). Assuming additive noise, the observation model writes:

$$\mathbf{y}_k^{obs} = \mathcal{H}_k(\mathbf{x}_k^t) + \boldsymbol{\varepsilon}_k^{obs}, \quad (3.31)$$

where $\boldsymbol{\varepsilon}_k^{obs}$ is the observation error at time t_k . $\mathcal{H}$ can represent the interpolation of model states to sensor locations, the application of physical measurement models, or any nonlinear transformation.

Observation errors arise from several sources, including measurement, processing, and representativeness errors (local-scale phenomena are captured in the observation but not accounted for in the model). The observation errors are supposed to be uncorrelated (with Gaussian errors, independent processes are uncorrelated) with background errors. The observation errors are considered as a realization of a Gaussian random variable:

$$\boldsymbol{\varepsilon}_k^{obs} \sim \mathcal{N}(0, \mathbf{R}_k), \quad (3.32)$$

where $\mathbf{R}_k$ is the observation error covariance matrix at time t_k . A common assumption is that $\mathbf{R}_k$ is diagonal, implying that the errors between observations are uncorrelated.

Despite the increasing availability of observations, from in-situ observations and remote sensing, the number of observations is usually lower than the number of model states in geosciences, with $N_{obs} \ll N_x$. Observations are used to improve the accuracy of the model; however, the model remains essential for interpolating and extrapolating between sparse observations.

3.3.6 Analysis

The parameter estimation problem with DA is summarized as:

$$\text{Find } \boldsymbol{\theta} \text{ such that: } \begin{cases} \boldsymbol{\theta} \text{ is near } \boldsymbol{\theta}^b \\ \mathcal{G}(\boldsymbol{\theta}) \text{ is near } \mathbf{y}^{obs} \end{cases}. \quad (3.33)$$

where the result of this analysis is $\mathbf{x}^a = \mathcal{M}(\boldsymbol{\theta}^a)$.

3.4 Bayesian approach

Data assimilation can be formulated within a Bayesian framework to combine the information from the observations and background knowledge. Under Gaussian assumptions and linear models, this leads to the BLUE, providing an analytical expression for parameter estimation.

3.4.1 Bayesian formulation of DA

DA can be defined with different mathematical frameworks (control theory, estimation) [Asch et al., 2016]. The states and parameters, as well as the observations, are modeled using random variables and are associated with probability density functions denoted as

$p(\cdot)$. The Bayesian framework provides a natural approach for fusing observations with the simulated state of the model. The goal is to improve the quantification of the prior probability of the parameters $\boldsymbol{\theta}^b$ or states $\mathbf{x}^b$ (both linked through the numerical model) given observations $\mathbf{y}^{obs}$. Then, by applying Bayes' theorem [Bayes, 1763], the problem writes:

$$p(\boldsymbol{\theta}|\mathbf{y}^{obs}) = \frac{p(\mathbf{y}^{obs}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y}^{obs})}, \quad (3.34)$$

where $p(\boldsymbol{\theta}|\mathbf{y}^{obs})$ is the posterior distribution of $\boldsymbol{\theta}$ conditioned on the observations $\mathbf{y}^{obs}$. $p(\boldsymbol{\theta})$ is the background knowledge on the parameters, $p(\mathbf{y}^{obs})$ is the probability of the observations $\mathbf{y}^{obs}$ independently from $\boldsymbol{\theta}$ and the likelihood $p(\mathbf{y}^{obs}|\boldsymbol{\theta})$ represents the probability of $\mathbf{y}^{obs}$ conditioned on $\boldsymbol{\theta}$. The term $p(\mathbf{y}^{obs})$ is usually not computed as it intervenes only as a normalization term and does not depend on $\boldsymbol{\theta}$, such that:

$$p(\boldsymbol{\theta}|\mathbf{y}^{obs}) \propto p(\mathbf{y}^{obs}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (3.35)$$

Bayesian model calibration is typically solved with MCMC methods such as the Metropolis-Hastings algorithm [Hastings, 1970]. These techniques require thousands of simulations to sample the distributions and reach convergence of the posterior distribution, which can be costly for evaluating expensive models. With DA assimilation techniques, the goal is to integrate new observations when they are available. Under additional assumptions, the Bayesian problem can be solved in a more computationally friendly way. For example, analytical solutions such as the BLUE and KF are presented in Sections 3.4.2 and 3.5.1, ensemble methods based on Monte-Carlo sampling in Section 3.5, and variational approaches in Section 3.6.

3.4.2 Best Linear Unbiased Estimator

BLUE analysis provides an exact solution to the Minimum Variance Estimation problem with Bayesian formulation under different assumptions:

- The numerical model $\mathcal{M}$ and the observation operator $\mathcal{H}$ are linear.
- Model and observation errors are independent and unbiased.
- The analysis is a linear combination of the background and observations, such that $\boldsymbol{\theta}^a = \mathbf{L}\boldsymbol{\theta}^b + \mathbf{K}\mathbf{y}^{obs}$, with $\mathbb{E}[\boldsymbol{\theta}^a] = 0$ unbiased and having minimal variance $\mathbb{V}ar[\boldsymbol{\theta}^a]$.

The objective is to find $\mathbf{L}$ and $\mathbf{K}$ for estimating the analyzed parameter $\boldsymbol{\theta}^a$. A simple interpretation of the BLUE analysis in the scalar case assumes that $\mathbf{H} = \mathbf{1}$. Under these assumptions, the BLUE analysis simplifies to:

$$\boldsymbol{\theta}^a = \boldsymbol{\theta}^b + \frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^{obs})^2}(\mathbf{y}^{obs} - \boldsymbol{\theta}^b), \quad (3.36)$$

$$\mathbb{V}ar[\boldsymbol{\theta}^a] = \frac{1}{(\sigma^b)^2} + \frac{1}{(\sigma^{obs})^2}, \quad (3.37)$$

where σ^b and σ^{obs} are the standard deviations of the background and observation errors, $\frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^{obs})^2}$ is called gain factor and $(\mathbf{y}^{obs} - \boldsymbol{\theta}^b)$ the innovation. In the general multidimensional case, without assuming $\mathbf{H} = \mathbf{1}$, the BLUE analysis writes:

$$\begin{cases} \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\ \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \\ \boldsymbol{\theta}^a = \boldsymbol{\theta}^b + \mathbf{K}(\mathbf{y}^{obs} - \mathbf{H}\boldsymbol{\theta}^b) \end{cases}, \quad (3.38)$$

where $\mathbf{P}^a$ is the analysis error covariance matrix, $\mathbf{K}$ is the Kalman gain and $(\mathbf{y}^{obs} - \mathbf{H}\theta^b)$ is the innovation. The derivation of these equations can be found in [Asch et al. \[2016\]](#). In the linear Gaussian case, Equation (3.38) is the MAP solution, providing an analytical solution to the Bayesian problem.

3.5 Bayesian approach for time-dependent models

In the BLUE analysis, the temporal evolution of the state variables is not considered. This formulation can be extended to temporal problems with similar ideas to assimilate observations sequentially, leading to the famous KF formulation. Then, alternative algorithms based on Kalman filtering are introduced, such as the EnKF or EKF. The derivation of the KF is presented as it is the basis of many DA approaches, such as the EnKF, which is later used in the manuscript, in Chapter 7.

3.5.1 Kalman Filter

The KF [[Kalman, 1960](#)] and the Kalman-Bucy filter [[Kalman and Bucy, 1961](#)] were introduced in the early 60s and initially applied to navigation for the Apollo Project to estimate the trajectory of the spacecraft. The KF is now widely used in various fields, including navigation, control systems, signal processing, robotics, finance, etc. This filter extends the BLUE analysis to temporal problems, using the same assumptions. In the KF, the control variable is usually the initial state of $\mathbf{x}$, but may concern both the initial state and/or parameters.

At each observation time t_k , the filter corrects the background estimate, referred to as the forecast $\mathbf{x}_k^f$, with the observation $\mathbf{y}_k^{obs}$. This correction produces a compromise analysis state $\mathbf{x}_k^a$ as shown in Figure 3.5.

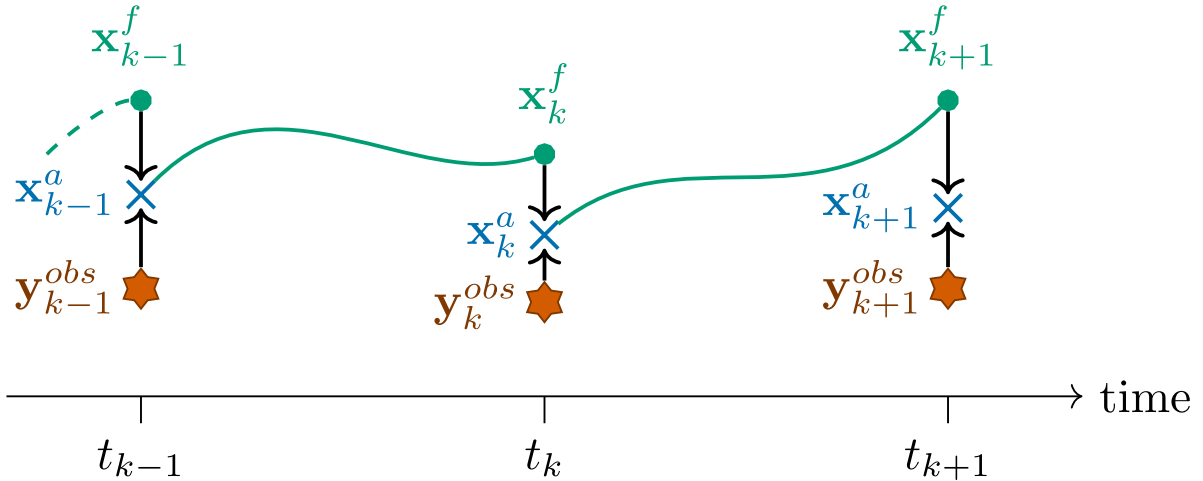


Figure 3.5: Sequential Data Assimilation principle with a Kalman Filter. The analyzed state $\mathbf{x}_k^a$ is corrected with $\mathbf{x}_k^f$ and observation $\mathbf{y}_k^{obs}$. The forecast $\mathbf{x}_k^f$ comes from model propagation starting from $\mathbf{x}_{k-1}^a$.

The following dynamic of the system is considered:

$$\begin{cases} \mathbf{x}_0^t = \mathbf{x}_0^b + \boldsymbol{\varepsilon}^b \\ \mathbf{x}_k^t = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1}^t + \boldsymbol{\eta}_k & \forall k \geq 1 \\ \mathbf{y}_k^{obs} = \mathbf{H}_k \mathbf{x}_k^t + \boldsymbol{\varepsilon}_k^{obs} & \forall k \geq 0 \end{cases} \quad (3.39)$$

where $\mathbf{M}_{k-1,k}$ is the transition matrix from time t_{k-1} to t_k . The KF proceeds in two alternative steps: forecasting and analysis.

Forecast step The forecast step propagates the model, as visualized by the green lines in Figure 3.5. The associated forecast error for the uncorrelated model and analysis errors is propagated with the covariance matrix:

$$\begin{aligned}\mathbf{P}_k^f &= \mathbb{E}[(\mathbf{x}_k^f - \mathbf{x}_k^t)(\mathbf{x}_k^f - \mathbf{x}_k^t)^T] \\ &= \mathbb{E}[(\mathbf{M}_{k-1,k}\boldsymbol{\varepsilon}_{k-1}^a - \boldsymbol{\eta}_k)(\mathbf{M}_{k-1,k}\boldsymbol{\varepsilon}_{k-1}^a - \boldsymbol{\eta}_k)^T] \\ &= \mathbf{M}_{k-1,k}\mathbb{E}[(\boldsymbol{\varepsilon}_{k-1}^a)(\boldsymbol{\varepsilon}_{k-1}^a)^T]\mathbf{M}_{k-1,k}^T + \mathbb{E}[(\boldsymbol{\eta}_{k-1})(\boldsymbol{\eta}_{k-1})^T] \\ &= \mathbf{M}_{k-1,k}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1,k}^T + \mathbf{Q}_k\end{aligned}\tag{3.40}$$

where $\boldsymbol{\varepsilon}_k^a = \mathbf{x}_k^a - \mathbf{x}_k^t$ is the analysis error, and $\mathbf{P}_k^a$ is the analysis error covariance matrix.

Analysis step Similarly to BLUE estimation, the forecast $\mathbf{x}_k^f$ is combined with the observation $\mathbf{y}_k^{obs}$ using a linear update:

$$\mathbf{x}_k^a = \mathbf{L}_k\mathbf{x}_k^f + \mathbf{K}_k\mathbf{y}_k^{obs},\tag{3.41}$$

The analysis error is defined as:

$$\boldsymbol{\varepsilon}_k^a = \mathbf{L}_k\boldsymbol{\varepsilon}_k^f + \mathbf{K}_k\boldsymbol{\varepsilon}_k^{obs} + (\mathbf{L}_k + \mathbf{K}_k\mathbf{H}_k - \mathbf{I})\mathbf{x}_k^t,\tag{3.42}$$

with $\mathbf{L}_k = \mathbf{I} - \mathbf{K}_k\mathbf{H}_k$ to ensure $\mathbb{V}ar[\boldsymbol{\varepsilon}^a] = 0$ is unbiased. Then, the Kalman gain $\mathbf{K}_k$ and the analysis error covariance matrix $\mathbf{P}_k^a$ write as:

$$\begin{aligned}\boldsymbol{\varepsilon}_k^a &= \mathbf{x}_k^a - \mathbf{x}_k^t \\ &= (\mathbf{x}_k^f - \mathbf{x}_k^t) + \mathbf{K}_k(\mathbf{y}_k^{obs} - \mathbf{H}_k\mathbf{x}_k^f) \\ &= \boldsymbol{\varepsilon}_k^f + \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k^t + \boldsymbol{\varepsilon}_k^{obs} - \mathbf{H}_k\mathbf{x}_k^f) \\ &= \boldsymbol{\varepsilon}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k) + \mathbf{K}_k\boldsymbol{\varepsilon}_k^{obs}\end{aligned}\tag{3.43}$$

$$\begin{aligned}\mathbf{P}_k^a &= \mathbb{E}[\boldsymbol{\varepsilon}_k^a(\boldsymbol{\varepsilon}_k^a)^T] \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T.\end{aligned}\tag{3.44}$$

The Kalman gain is computed by minimizing the trace of the analysis error covariance matrix (need to compute the derivative of the trace with respect to $\mathbf{K}_k$), leading to:

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{R}_k + \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T)^{-1}.\tag{3.45}$$

The Kalman gain $\mathbf{K}_k$ is a balance between the background and the observations, weighted by their respective uncertainties.

Summary

$$\begin{aligned}\text{Forecast step} \quad & \mathbf{x}_k^f = \mathbf{M}_{k-1,k}\mathbf{x}_{k-1}^a \\ & \mathbf{P}_k^f = \mathbf{M}_{k-1,k}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1,k}^T + \mathbf{Q}_k \\ \text{Analysis step} \quad & \mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{R}_k + \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T)^{-1} \\ & \mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^{obs} - \mathbf{H}_k\mathbf{x}_k^f) \\ & \mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T\end{aligned}\tag{3.46}$$

The KF is the optimal solution of $p(\mathbf{x}_k | \mathbf{y}_{k:0}^{obs})$ for Gaussian errors and linear models.

Remark: Environmental models are often nonlinear, which limits the applicability of KF. Furthermore, the storage of covariance matrices and matrix computations can be very costly in high dimensions. Therefore, the KF is rarely used in geosciences. Instead, approximations of the KF are used. For instance, the model can be replaced with TLM, leading to the EKF formulation. Another alternative is the EnKF, which addresses nonlinearity and computational cost by representing error statistics using an ensemble of model states.

3.5.2 Extended Kalman Filter

The EKF is an adaptation of the KF designed to handle nonlinearities in both the model and observation operators. The EKF principle is based on a local linearization of the operators at each time step around the current estimate of the state [Jazwinski, 1970]. The nonlinear model propagates the mean state, but the error covariance evolution is approximated using tangent linear operators. The tangent linear operators near the forecast state are obtained with a Taylor series expansion:

$$\begin{aligned}\mathcal{M}_{k,k+1}(\mathbf{x}_k^a) &= \mathcal{M}_{k,k+1}(\mathbf{x}_k^f) + \mathbf{M}_{k,k+1} \left(\mathbf{x}_k^a - \mathbf{x}_k^f \right) + \mathcal{O} \left(\|\mathbf{x}_k^a - \mathbf{x}_k^f\|^2 \right), \\ \mathcal{H}_k(\mathbf{x}_k^a) &= \mathcal{H}_k(\mathbf{x}_k^f) + \mathbf{H}_k \left(\mathbf{x}_k^a - \mathbf{x}_k^f \right) + \mathcal{O} \left(\|\mathbf{x}_k^a - \mathbf{x}_k^f\|^2 \right).\end{aligned}\tag{3.47}$$

where $\mathbf{M}_{k,k+1}$ and $\mathbf{H}_k$ are the linear tangent operators (Jacobian matrices) evaluated at the forecast state $\mathbf{x}_k^f$ of the model and observation operators, respectively. The EKF also assumes that the errors follow Gaussian distributions.

For highly nonlinear operators, the analysis produced by the EKF is not optimal, and its accuracy depends on the degree of nonlinearity [Miller et al., 1994]. The Gaussian error covariance may deviate from a Gaussian distribution, and the EKF approximates the estimation error distributions [Carrassi et al., 2018]. While the EKF was successfully applied in meteorology [Ghil et al., 1981] and oceanography [Ghil and Malanotte-Rizzoli, 1991] studies, its application to large-scale dynamics is limited. The computation and storage of the Kalman gain $\mathbf{K}$ and the error covariance matrices remain challenging for high-dimensional geophysical models, such as Numerical Weather Prediction models or 2D flood models. Indeed, the error covariance matrix $\mathbf{P}$ and the Kalman gain $\mathbf{K}$ dimensions are proportional to the square of the state size. In geophysical systems such as Numerical Weather Prediction models or 2D flood models, the number of states generally ranges from 10^6 to 10^{10} variables. Hence, the storage and computation of the matrices are prohibitive and exceed the memory capacities of most computer architectures [Evensen, 2009a]. Thus, the EKF, similarly to the KF (described and derived in Section 3.5.1), can only be applied to low-dimensional systems. Therefore, ensemble-based approaches, such as the EnKF, were developed because they avoid the explicit storage of $\mathbf{K}$ and $\mathbf{P}$.

3.5.3 Ensemble Kalman Filter

The EnKF was developed in the 1990s by Evensen [1994] as an alternative to the KF and EKF for high-dimensional systems. Unlike traditional KF, covariance matrices are propagated with Monte Carlo sampling of model realizations. As illustrated in Figure 3.6, an ensemble (of members) of N forecasted states $\{(\mathbf{x}_k^f)^{(1)}, \dots, (\mathbf{x}_k^f)^{(N)}\}$ is considered resulting from the propagation of the nonlinear models applied to the analysis state ensemble

$\{(\mathbf{x}_{k-1}^a)^{(1)}, \dots, (\mathbf{x}_{k-1}^a)^{(N)}\}$ from the previous assimilation. Each member of this ensemble is independently updated using the classical KF. The gain and covariance matrices $\mathbf{P}_k^a$ and $\mathbf{P}_k^f$ are estimated empirically from the ensemble mean and covariance.

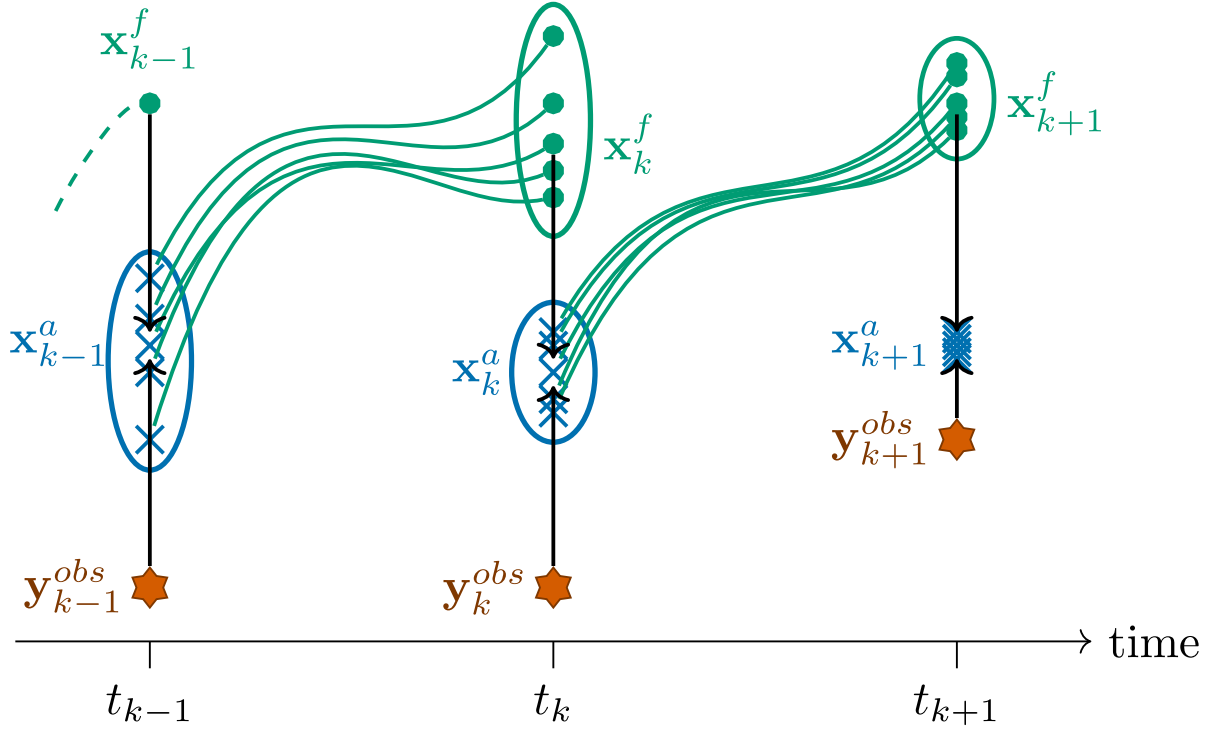


Figure 3.6: Sequential Data Assimilation with Ensemble Kalman Filter. Analysis samples are computed (blue crosses) based on the observation $\mathbf{y}^{obs}$ (orange hexagons) and the forecasted ensemble $\mathbf{x}^f$ (green dots). At each time step, the analysis ensemble is propagated by the model (green trajectories).

Like previous sequential methods, the EnKF consists of a forecast and analysis step. The following derivation, which considers a linear observation operator, can be easily adapted to nonlinear operators [Evensen, 2009b].

Forecast step The forecast step generates and propagates an ensemble of N members. The members are generated by sampling the distribution of the model parameters (the background) or by perturbing the initial forecast state, such as:

$$\forall i \in \{1, \dots, N\} \quad (\mathbf{x}_0^f)^{(i)} = \mathbf{x}_0^f + \delta^{(i)}, \quad (3.48)$$

where $\delta^{(i)}$ is a perturbation of the initial forecast. For updating the forecast covariance matrix, an ensemble-anomaly covariance [Carrassi et al., 2018] is defined as:

$$(\mathbf{E}_k^f)^e = \frac{1}{\sqrt{N-1}} [(\mathbf{x}_k^f)^{(1)} - \bar{\mathbf{x}}_k^f, \dots, (\mathbf{x}_k^f)^{(N)} - \bar{\mathbf{x}}_k^f], \quad (3.49)$$

where the ensemble mean is $\bar{\mathbf{x}}_k^f = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_k^f)^{(i)}$. The superscript e denotes ensemble-based estimates. Similarly, the ensemble-anomaly covariance matrix for the analysis is derived with superscript a in place of f . The ensemble-based error covariance matrices are then given by:

$$(\mathbf{P}_k^f)^e = ((\mathbf{E}_k^f)^e)((\mathbf{E}_k^f)^e)^T, \quad (3.50)$$

$$(\mathbf{P}_k^a)^e = ((\mathbf{E}_k^a)^e)((\mathbf{E}_k^a)^e)^T. \quad (3.51)$$

Analysis step The analysis step defines the Kalman gain $\mathbf{K}_k$, the analysis $\mathbf{x}_k^a$ and the analysis covariance matrix $\mathbf{P}_k^a$. According to Evensen [1994], updating each ensemble member with KF Equations (3.46) results in an ensemble with the correct mean and covariance matrix, and there is no need to compute the full covariance matrices explicitly. Furthermore, Burgers et al. [1998] proved that the EnKF analysis is consistent with the KF when the observations are treated as random variables. Then, we consider an ensemble of observations such as:

$$\forall i \in \{1, \dots, N\} \quad (\mathbf{y}_k^{obs})^{(i)} = \mathbf{y}_k^{obs} + (\boldsymbol{\varepsilon}_k^{obs})^{(i)}. \quad (3.52)$$

Similarly to the forecast step, an ensemble-anomaly matrix of the observations is considered, such that:

$$(\mathbf{D}_k^{obs})^e = \frac{1}{\sqrt{N-1}} [(\mathbf{y}_k^{obs})^{(1)} - \bar{\mathbf{y}}_k^{obs}, \dots, (\mathbf{y}_k^{obs})^{(N)} - \bar{\mathbf{y}}_k^{obs}], \quad (3.53)$$

where $\bar{\mathbf{y}}_k^{obs} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_k^{obs})^{(i)}$ is the mean of the ensemble of perturbed observations. Then the ensemble covariance matrix is defined as:

$$\mathbf{R}_k^e = (\mathbf{D}_k^{obs})^e ((\mathbf{D}_k^{obs})^e)^T. \quad (3.54)$$

As $N \rightarrow \infty$, this converges to the error covariance matrix defined for the KF. Each ensemble member is then updated as:

$$\forall i \in \{1, \dots, N\} \quad (\mathbf{x}_k^a)^{(i)} = (\mathbf{x}_k^f)^{(i)} + (\mathbf{P}_k^f)^e \mathbf{H}_k^T [\mathbf{H}_k (\mathbf{P}_k^f)^e \mathbf{H}_k^T + \mathbf{R}_k^e]^{-1} [(\mathbf{y}_k^{obs})^{(i)} - \mathbf{H}_k (\mathbf{x}_k^f)^{(i)}], \quad (3.55)$$

with analysis ensemble mean:

$$\bar{\mathbf{x}}_k^a = \bar{\mathbf{x}}_k^f + (\mathbf{P}_k^f)^e \mathbf{H}_k^T [\mathbf{H}_k (\mathbf{P}_k^f)^e \mathbf{H}_k^T + \mathbf{R}_k^e]^{-1} [\bar{\mathbf{y}}_k^{obs} - \mathbf{H}_k \bar{\mathbf{x}}_k^f]. \quad (3.56)$$

Finally, the analysis ensemble-anomaly matrix $(\mathbf{E}_k^a)^e$ is computed analogously to $(\mathbf{E}_k^f)^e$, and the analysis covariance matrix is:

$$\begin{aligned} (\mathbf{P}_k^a)^e &= ((\mathbf{E}_k^a)^e)((\mathbf{E}_k^a)^e)^T \\ &= (\mathbf{I} - \mathbf{K}_k^e \mathbf{H}_k)(\mathbf{P}_k^f)^e \end{aligned} \quad (3.57)$$

with $\mathbf{K}_k^e = (\mathbf{P}_k^f)^e \mathbf{H}_k^T [\mathbf{H}_k (\mathbf{P}_k^f)^e \mathbf{H}_k^T + \mathbf{R}_k^e]^{-1}$.

Summary

$$\begin{aligned} \text{Forecast step} \quad \forall i \in \{1, \dots, N\} \quad & (\mathbf{x}_k^f)^{(i)} = \mathbf{M}_{k-1,k}(\mathbf{x}_{k-1}^a)^{(i)} \\ & (\mathbf{E}_k^f)^e = \frac{1}{\sqrt{N-1}} [(\mathbf{x}_k^f)^{(1)} - \bar{\mathbf{x}}_k^f, \dots, (\mathbf{x}_k^f)^{(N)} - \bar{\mathbf{x}}_k^f] \\ & (\mathbf{P}_k^f)^e = ((\mathbf{E}_k^f)^e)((\mathbf{E}_k^f)^e)^T \\ \text{Analysis step} \quad & \mathbf{K}_k^e = (\mathbf{P}_k^f)^e \mathbf{H}_k^T [\mathbf{H}_k (\mathbf{P}_k^f)^e \mathbf{H}_k^T + \mathbf{R}_k^e]^{-1} \\ \forall i \in \{1, \dots, N\} \quad & (\mathbf{x}_k^a)^{(i)} = (\mathbf{x}_k^f)^{(i)} + \mathbf{K}_k^e ((\mathbf{y}_k^{obs})^{(i)} - \mathbf{H}_k (\mathbf{x}_k^f)^{(i)}) \\ & (\mathbf{P}_k^a)^e = (\mathbf{I} - \mathbf{K}_k^e \mathbf{H}_k)(\mathbf{P}_k^f)^e \end{aligned} \quad (3.58)$$

EnKF is well suited for high-dimensional systems because it avoids the computation and storage of large covariance matrices explicitly. It is currently the most widely used algorithm for DA [Carrassi et al., 2018]. In flood modeling, it is often used to update the water depths in the model for near-real-time flood mapping by assimilating flood extent or water depth observations [Neal et al., 2007; Ziliani et al., 2019; Jafarzadegan et al., 2021].

3.5.4 Particle Filter

The PF solves the Bayes Equation (3.34) based on a sequential Monte Carlo method, but without assuming a linear model or Gaussian error assumptions. It approximates the posterior distribution probability with random samples (called particles). Each particle represents a possible state of the system, and the ensemble of particles provides a Monte Carlo approximation conditioned on the observations. The simplest version is the bootstrap PF introduced in Gordon et al. [1993]. The forecast probability density $p(\mathbf{x}_k | \mathbf{y}_{0:k-1}^{obs})$ is approximated by a weighted empirical distribution with particles such as:

$$p(\mathbf{x}_k | \mathbf{y}_{0:k-1}^{obs}) \approx \sum_{i=1}^N \omega_{k-1}^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}), \quad (3.59)$$

where $\{\mathbf{x}_k^{(i)}\}_{i=1}^N$ is the particles, $\{\omega_{k-1}^{(i)}\}_{i=1}^N$ is their associated weights, and δ is delta-Dirac masses each centered on a state $\mathbf{x}_k^{(i)}$. The weights are normalized to sum to 1, since Equation (3.59) represents a probability density function. A usual assumption for the initialization at $t = 0$ is uniform weights. This approximation is the main simplification introduced by the PF when using a finite ensemble N .

Forecast step The forecast step propagates the N particles forward in time with the model:

$$\begin{aligned} p(\mathbf{x}_{k+1} | \mathbf{y}_{0:k}^{obs}) &= \sum_{i=1}^N \omega_k^{(i)} p(\mathbf{x}_{k+1} | \mathbf{x}_k^{(i)}) \\ &\propto \sum_{i=1}^N \omega_k^{(i)} \delta(\mathbf{x}_{k+1} - \mathbf{x}_k^{(i)}). \end{aligned} \quad (3.60)$$

Analysis step The analysis state consists of assimilating observation at time t_k using the Bayes rule:

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{y}_{0:k}^{obs}) &= \frac{p(\mathbf{y}_k^{obs} | \mathbf{x}_k)}{p(\mathbf{y}_{0:k}^{obs})} p(\mathbf{x}_k | \mathbf{y}_{0:k-1}^{obs}) \\ &= \sum_{i=1}^N \frac{\omega_{k-1}^{(i)} p(\mathbf{y}_k^{obs} | \mathbf{x}_k^{(i)})}{p(\mathbf{y}_{0:k}^{obs})} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}) \\ &\propto \sum_{i=1}^N \omega_{k-1}^{(i)} p(\mathbf{y}_k^{obs} | \mathbf{x}_k^{(i)}) \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}). \end{aligned} \quad (3.61)$$

The update step suggests that the particle weights are updated:

$$\omega_k^{(i)} \propto \omega_{k-1}^{(i)} p(\mathbf{y}_k^{obs} | \mathbf{x}_k^{(i)}). \quad (3.62)$$

where $p(\mathbf{y}_k^{obs} | \mathbf{x}_k^{(i)})$ is the likelihood.

Degeneracy and resampling A well-known problem of the PF is the degeneracy problem, where after several assimilation steps, one or a few particles largely dominate the others, leading to poor distributions characterization of distributions. To mitigate this problem, a resampling step is often introduced after the analysis step and before the next forecasting step. Resampling generates a new set of particles $\{\mathbf{x}_k^{(i)}\}_{1 \leq i \leq N}$ with replacement, according to their weights. There are several methods for resampling the particles,

including multinomial, stratified, residual, and systematic resampling. Once resampled, all the particles are usually assigned equal weights in the classical SIR algorithm.

Remark: The PF is interesting due to its minimal assumptions, but it suffers from the curse of dimensionality. The number of particles required to represent the posterior probability distribution increases exponentially with the dimension of the state space. Thus, its application is usually limited to low-dimensional problems, typically with a state dimension of 10 to 20. For instance, they have been successfully applied to low-dimensional problems such as target oceanographic models with simplified dynamics [Van Leeuwen, 2009].

Note: In Chapter 7, the EnKF and PF are used to estimate the spatially variable roughness on a schematic test case when assimilating water depth observations. As suggested by this synthetic review of the methods, the state dimension should remain limited to apply it. The PF is particularly costly, and the EnKF requires a large ensemble to approximate the error covariance, which is computationally expensive with our numerical model. Thus, variational approaches were mainly used in the manuscript with a reduced computational cost and fewer tuning parameters (no ensemble size). The variational approaches are presented in Section 3.6.

3.6 Variational approaches

Variational approaches were historically introduced by Sasaki [1955, 1958] and extended in the 1980s by Le Dimet and Talagrand [1986] and Talagrand and Courtier [1987] using optimal control theory [Lions and Lelong, 1968]. These approaches are designed to estimate the best trajectory of the model to fit to observations within an assimilation window $[t_0, t_K]$ (for oceanography or flood modeling, the assimilation window can vary from some hours to some days). The initial background state $\mathbf{x}_0^b$ or the model parameters $\boldsymbol{\theta}_0^b$ are changed gradually to get the closest trajectory of the model states to the observations within the assimilation window. In practice, the unknowns are evaluated by quantifying the misfit between model predictions and measurements by minimizing a cost function.

3.6.1 Variational formulation

Over an assimilation window $[t_0, t_K]$, if an observation is available at each time step an extended state vector containing all the time steps is $\mathbf{x} = [\mathbf{x}_0, \dots, \mathbf{x}_K]$. The complete variational DA scheme formulates as the minimization of a least square cost function in terms of their covariance matrices $\mathbf{B}$, $\mathbf{R}$, and $\mathbf{Q}$, over an assimilation window $[t_0, t_K]$:

$$J(\mathbf{x}) = \underbrace{\|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2}_{J_b} + \underbrace{\|\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2}_{J_o} + \underbrace{\|\mathcal{F}(\mathbf{x})\|_{\mathbf{Q}^{-1}}^2}_{J_q}, \quad (3.63)$$

where $\mathcal{F}_k(\mathbf{x}) = \mathbf{x}_k - \mathcal{M}(\mathbf{x}_{k-1})$ and $\|\mathbf{x}\|_A^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$ is the Euclidean norm with a weighting matrix. With this minimization problem, the state trajectory is estimated not far from the background (J_b), from the observation (J_o), and following the model equations (J_q). In Equation (3.63), model linearity is not assumed, and no assumptions are made on the statistical nature of model errors. This minimization problem is complex due to the dimension of $\mathbf{x}$ (the size of the state vector multiplied by the number of time steps in the assimilation window), and model errors are generally difficult to define. A usual

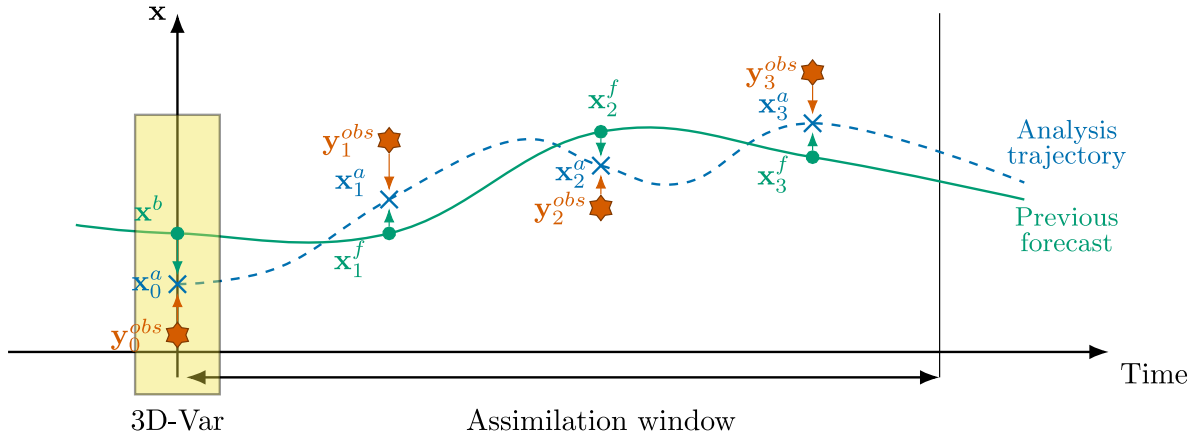


Figure 3.7: Three-Dimensional Variational (3D-Var) and 4D-Var assimilation scheme representation. The aim is to determine the optimal analysis trajectory (dashed blue curve) from a background trajectory (green curve) and a set of observations (orange marks).

assumption is to neglect model errors. The problem is said to be a “strong constraint” with this additional hypothesis, as opposed to “weak constraint” problems, including model errors. In practice, additional hypotheses and approximations are considered to implement variational DA. Three approximations are presented in the following sections: the 3D-Var, the 4D-Var and the Incremental 4D-Var formulations.

3.6.2 Three-Dimensional Variational formulation

The 3D-Var algorithm is a non-temporal problem that minimizes a cost function to find the best compromise between the observations and the background, as visualized in Figure 3.7. Its formulation simplifies the problem defined in Equation (3.63) considering that the model is the identity ($\mathcal{M} = \mathbf{I}$) over the assimilation window. The problem is defined only by its initial state, resulting in $\mathbf{x} = \mathbf{x}_0$ and $J_q = 0$. The cost function to minimize is then defined as:

$$\begin{aligned} J(\mathbf{x}_0) &= (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + (\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x}_0))^T \mathbf{R}^{-1} (\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x}_0)) \\ &= J_b(\mathbf{x}_0^b) + J_o(\mathbf{x}_0). \end{aligned} \quad (3.64)$$

Three-Dimensional Variational formulation and MAP The 3D-Var formulation is closely linked to the evaluation of the MAP of the Bayes theorem to determine $\mathbf{x}_0$ conditioned on the observations. We recall the Bayes formulation formulated in Equation (3.34)

$$p(\mathbf{x}_0 | \mathbf{y}^{obs}) = \frac{p(\mathbf{y}^{obs} | \mathbf{x}_0) p(\mathbf{x}_0)}{p(\mathbf{y}^{obs})}. \quad (3.65)$$

The marginal likelihood on the observations $p(\mathbf{y}^{obs})$ does not depend on $\mathbf{x}_0$ so the computation of the MAP is equivalent to maximizing:

$$p(\mathbf{x}_0 | \mathbf{y}^{obs}) \propto p(\mathbf{y}^{obs} | \mathbf{x}_0) p(\mathbf{x}_0). \quad (3.66)$$

The errors on the control parameters $\mathbf{x}_0$ and on the observations $\mathbf{y}^{obs}$ are considered random variables defined by Gaussian distributions, then:

$$p(\mathbf{x}_0) = \frac{1}{(2\pi)^{N_x/2} \det(\mathbf{B})^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) \right). \quad (3.67)$$

$$p(\mathbf{y}^{obs} | \mathbf{x}_0) = \frac{1}{(2\pi)^{N_{obs}/2} \det(\mathbf{R})^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x}_0))^T \mathbf{R}^{-1} (\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x}_0)) \right) \quad (3.68)$$

Finally, by taking the logarithm of Equation (3.66) and removing the constant non-dependent on $\mathbf{x}_0$:

$$\begin{aligned} \ln p(\mathbf{x}_0 | \mathbf{y}^{obs}) &\propto \ln(p(\mathbf{y}^{obs} | \mathbf{x}_0) p(\mathbf{x}_0)) \\ &= (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) \\ &\quad + (\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x}_0))^T \mathbf{R}^{-1} (\mathbf{y}^{obs} - \mathcal{H}(\mathbf{x}_0)) \\ &= J_b(\mathbf{x}_0^b) + J_o(\mathbf{x}_0). \end{aligned} \quad (3.69)$$

Thus, minimizing $J(\mathbf{x}_0)$ is equivalent to maximizing the a posteriori probability density.

3D-Var and BLUE analysis For a linear observation operator, the solution of the BLUE is equivalent to the 3D-Var formulation with the cost function J having a unique optimum $\mathbf{x}_0^a$. The 3D-Var approaches the optimal with gradient descent in the direction:

$$\nabla J(\mathbf{x}_0) = 2\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - 2\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}^{obs} - \mathbf{H}\mathbf{x}_0) \quad (3.70)$$

At the minimum of the cost function, $\nabla J(\mathbf{x}_0) = 0$, so:

$$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1}) \mathbf{x}_0^a = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^{obs} + \mathbf{B}^{-1} \mathbf{x}_0^b \quad (3.71)$$

Rewriting this, after several steps, the analysis update is:

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \mathbf{K}(\mathbf{y}^{obs} - \mathbf{H}\mathbf{x}_0^b) \quad (3.72)$$

where the Kalman gain matrix $\mathbf{K}$ is defined as:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (3.73)$$

This is the solution of the BLUE, proving the equivalence between the 3D-Var cost function minimization and the BLUE estimator for linear observation operators and Gaussian errors.

3.6.3 Four-Dimensional Variational formulation

As opposed to 3D-Var, 4D-Var distinguishes the temporal dimensions for time-dependent problems. While 3D-Var assimilates all observations at a single time step, 4D-Var incorporates observations distributed across multiple time steps.

As in 3D-Var, 4D-Var considers a strong constraint problem, meaning that the model is perfect and error-free. Thus, the model error term is omitted in the cost function ($J_q = 0$), and the model state trajectory $\mathbf{x}$ is fully determined from the initial state $\mathbf{x}_0$. For determining this initial state, $\mathcal{H}_k(\mathbf{x}_k)$ is compared with the observations $\mathbf{y}_k^{obs}$ at various times t_k as shown in Figure 3.7. The cost function balances the fit to the prior background trajectory and agreement with observations:

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0). \quad (3.74)$$

The discrepancy between $\mathbf{x}_0$ and the background term is similar to 3D-Var:

$$J_b(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b). \quad (3.75)$$

The observation term accounts for discrepancies between model predictions and observations at all observation times:

$$J_0(\mathbf{x}_0) = \sum_{k=0}^K (\mathbf{y}_k^{obs} - \mathcal{H}_k(\mathbf{x}_k))^T \mathbf{R}_k^{-1} (\mathbf{y}_k^{obs} - \mathcal{H}_k(\mathbf{x}_k)). \quad (3.76)$$

Given the deterministic nature of the model, each state $\mathbf{x}_k$ can be expressed recursively as a function of the initial state $\mathbf{x}_0$:

$$\begin{aligned} \mathbf{x}_k &= \mathcal{M}_{k:0}(\mathbf{x}_0) \\ &= \mathcal{M}_{k-1,k} \mathcal{M}_{k-2,k-1} \dots \mathcal{M}_{1,2} \mathcal{M}_{0,1} \mathbf{x}_0 \\ &= \mathcal{M}_k \circ \mathcal{M}_{k-1} \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0). \end{aligned} \quad (3.77)$$

Finally, the 4D-Var cost function is defined as:

$$\begin{aligned} J(\mathbf{x}_0) &= J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) \\ &= (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \\ &\quad \sum_{k=0}^K (\mathbf{y}_k^{obs} - \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0)))^T \mathbf{R}_k^{-1} (\mathbf{y}_k^{obs} - \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0))). \end{aligned} \quad (3.78)$$

Remark: For linear operators $\mathcal{M}$ and $\mathcal{H}$, and with KF hypotheses on the error covariance matrices, the minimization of the cost function at the end of the assimilation window results in the same solution as the final analysis of the KF.

The 4D-Var formulation can be further simplified by considering that $\mathcal{H}$ is linear, and by using the Incremental 4D-Var or Multi-Incremental 4D-Var algorithms, which linearize the dynamics around the current trajectory and minimize a convex cost function.

3.6.4 Incremental Four-Dimensional Variational formulation

The 4D-Var assimilation cost function defined in Equation (3.78) involves a nonlinear numerical model. The optimization procedure may result in local minima. To address this issue, [Courtier et al. \[1994\]](#) introduced the Incremental 4D-Var, which enables the minimization of a quadratic convex cost function. For weakly nonlinear models and nonlinear observation operators, the dynamics can be linearized around the current trajectory, similar to the EKF described in Section 3.5.2. The goal is to minimize a convex cost function for an increment $\delta \mathbf{x}_0$:

$$\delta \mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^b, \quad (3.79)$$

where $\mathbf{M}_k$ and $\mathbf{H}_k$ are the linearizations of the model and observation operators, respectively, such as:

$$\begin{aligned} \mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0) - \mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0^b) &\approx \mathbf{M}_k \dots \mathbf{M}_1(\mathbf{x}_0 - \mathbf{x}_0^b) \\ \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0)) - \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0^b)) &\approx \mathbf{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0) - \mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0^b)). \end{aligned} \quad (3.80)$$

With this linearization formulation, the cost function writes with respect to this increment:

$$\begin{aligned}
 J(\delta \mathbf{x}_0) &= J_b(\delta \mathbf{x}_0) + J_o(\delta \mathbf{x}_0) \\
 J_b(\delta \mathbf{x}_0) &= \delta \mathbf{x}_0^T \mathbf{B}^{-1} \delta \mathbf{x}_0 \\
 J_o(\delta \mathbf{x}_0) &= \sum_{k=0}^K (\mathbf{y}_k^{obs} - \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\delta \mathbf{x}_0)))^T \mathbf{R}_k^{-1} (\mathbf{y}_k^{obs} - \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\delta \mathbf{x}_0))) \\
 &= \sum_{k=0}^K (\mathbf{d}_k - \mathbf{H}_k \mathbf{M}_k \dots \mathbf{M}_1 \delta \mathbf{x}_0)^T \mathbf{R}_k^{-1} (\mathbf{d}_k - \mathbf{H}_k \mathbf{M}_k \dots \mathbf{M}_1 \delta \mathbf{x}_0).
 \end{aligned} \tag{3.81}$$

where $\mathbf{d}_k = \mathbf{y}_k^{obs} - \mathcal{H}_k(\mathcal{M}_k \circ \dots \circ \mathcal{M}_1(\mathbf{x}_0^b))$ is the innovation vector. The resulting incremental cost function $J(\delta \mathbf{x}_0)$ is a quadratic approximation of the initial cost function. It results in a unique absolute minimum, and it can be minimized more easily. The non-linearities are treated incrementally since the linearized models are updated for each new trajectory.

Note: In the present thesis, we have chosen to use 3D-Var formulation which is applied in Chapter 6 and 7. 3D-Var was chosen as a cheaper alternative to avoid model and adjoint integration of the 4D-Var. While 3D-Var was designed for static problems, it can be used for high-dimensional dynamical problems. All the observations in an assimilation window $[t_0, t_K]$ are assimilated as if observed at the initial time $t = 0$. Practically, Equation (3.64) is minimized, and all of the observations are concatenated in a single vector $\mathbf{y}^{obs}$, with the operator $\mathcal{H}$ mapping the model states into the observation space (here the space with all the observations). The error covariance matrix $\mathbf{R}$ is adapted to account for the uncertainties of all the observations.

3.6.5 Error covariance matrices: $\mathbf{B}$ and $\mathbf{R}$

In variational DA, the estimation of the state variables depends on the definition of errors $\mathbf{B}$ and $\mathbf{R}$. As in the KF algorithm, when the background covariance matrix $\mathbf{B}$ is high-dimensional, its inversion and storage are computationally challenging. The observation error covariance matrix $\mathbf{R}$ is usually smaller ($N_{obs} \ll N_x$), but it can also become large and difficult to handle.

Background covariance matrix $\mathbf{B}$ For the KF storing $\mathbf{B}$ is impossible for high-dimensional systems. However, in the variational approach, only matrix-vector products involving the inverse $\mathbf{B}^{-1}$ are computed, which avoids the need to store or explicitly invert the full matrix. Instead of constructing the matrix, one defines an operator that takes a vector $\mathbf{A}$ as input and returns the product $\mathbf{B}^{-1}\mathbf{A}$ as output, as described in [Weaver and Courtier \[2001\]](#) and [Ricci \[2004\]](#).

Observation covariance matrix $\mathbf{R}$ The observation errors are usually considered to be uncorrelated in both space and time [[Cao et al., 2007](#)], leading to a diagonal $\mathbf{R}$, which is trivial to invert. However, this assumption is often violated in environment models (e.g., in a 2D flood model, the errors on water depth measurements between two neighboring positions are usually correlated in time and space). Another possibility is to define a parametric correlation length between the errors, resulting in a sparse covariance matrix with non-zero entries along the diagonal and near diagonal (e.g., spatial measurement of the same variable at neighboring locations or times). In this approach, the covariance

between two observation errors is modeled as a function of the spatial or temporal distance between the observations, with, for instance, an exponential kernel such that:

$$\mathbf{R}_{i,j} = (\sigma^{obs})^2 \exp\left(\frac{-d_{i,j}}{L}\right) \quad (3.82)$$

where L is a defined correlation length which determines the strength of the spatial error correlation, $d_{i,j}$ is the distance between the observation points (e.g., Euclidean distance between the points locations) and $(\sigma^{obs})^2$ is the error variance. Another standard structure is block diagonal covariance matrices, where independent variables are stored in $\mathbf{y}^{obs}$ and each of them is characterized by dependent errors, as in Robert et al. [2005]. These hypotheses rarely hold completely in environmental applications [Robert et al., 2005]. As an alternative, $\mathbf{R}$ can be estimated empirically from ensembles of repeated observations. If a physical quantity is observed n times, the deviations from the sample mean can be stored in a vector $\mathbf{A}'$, and the covariance matrix defined as $\mathbf{R} = \frac{1}{n-1} \mathbf{A}'(\mathbf{A}')^T$.

Remark: Quantifying the errors $\mathbf{B}$ and $\mathbf{R}$ is key to finding the compromise between the background and the observations. In some calibration studies, observations are assumed to be nearly error-free, resulting in small observation errors relative to the background. However, when observation errors are non-negligible, $\mathbf{B}$ and $\mathbf{R}$ should be carefully estimated to reach the right balance between the background and the observations. In Chapters 6 and 7, the role of the observation errors on the characterization of the roughness field is evaluated.

3.6.6 Solutions to the variational problems

The minimization of the cost functions defined for the previous variational approaches is the following problem:

$$J(\mathbf{x}_0) = \min_{\mathbf{x}_0 \in \mathbb{R}^{N_x}} J(\mathbf{x}_0), \quad (3.83)$$

The minimization methods based on gradients are usually described as:

$$\mathbf{x}_0^{i+1} = \mathbf{x}_0^i + \Delta^i \mathbf{d}^i \quad (3.84)$$

where $\mathbf{d}^i \in \mathbb{R}^{N_x}$ is the descent direction at iteration i , and $\Delta^i \in \mathbb{R}$ is the step size at iteration i . Different descent methods vary in how the direction $\mathbf{d}^i$ and step size Δ^i are computed, such as steepest descent, Newton, or quasi-Newton methods. The cost function value should decrease with the number of iterations, and $\mathbf{x}_0^i$ tends towards the minimum more or less quickly depending on the descent method. In these strategies, the gradient of the cost function is an essential component, which can be either computed with finite difference schemes or using the adjoint of the model. The following subsections describe famous descent methods, and the finite difference and adjoint method techniques to compute the gradients.

3.6.6.1 Descent methods

One of the simplest descent methods is the steepest descent method. The steepest method aims to find the minimum with a sequence $\mathbf{x}_0^i$ such that $J(\mathbf{x}_0^i)$ decreases with the iteration number i until the minimum is reached. The steepest direction is identified based on the negative gradient $\mathbf{d}^i = -\nabla J(\mathbf{x}_0^i)$ and the step size Δ^i via line search. For the line search, once the direction $\mathbf{d}^i$ for iteration i is identified, Δ^i can be quantified via minimization of

$f(\Delta) = J(\mathbf{x}_0^i + \Delta^i \mathbf{d}^i)$. It is a function of a single variable, so it is easy to minimize in a few iterations with methods such as the one introduced in [Powell \[1964\]](#).

However, the steepest slope direction is not always the best direction, and alternative methods to converge faster to the minimum have been proposed, such as methods based on the Newton and Quasi-Newton methods. The Newton method accounts for the curvature of the cost function via the Hessian matrix, such as:

$$\mathbf{x}_0^{i+1} = \mathbf{x}_0^i - [\nabla^2 J(\mathbf{x}_0^i)]^{-1} \nabla J(\mathbf{x}_0^i), \quad (3.85)$$

where $\nabla^2 J(\mathbf{x}_0^i)$ is the Hessian of J , assuming its invertibility. The computation of the Hessian at each iteration is prohibitive for large-scale problems.

To overcome this limitation, quasi-Newton methods were developed. These algorithms construct an approximation of the Hessian or its inverse only based on gradient evaluation (ensuring a limited computation cost). Among the most widely used quasi-Newton methods is the BFGS algorithm [[Broyden, 1970](#); [Fletcher, 1970](#); [Goldfarb, 1970](#); [Shanno, 1970](#)], which updates the inverse Hessian approximation using gradient differences and step differences from previous iterations. The L-BFGS [[Liu and Nocedal, 1989](#)] or its bounded L-BFGS-B version [[Byrd et al., 1995](#)], which imposes bounds on the parameters to optimize, are particularly well-suited for large-dimensional problems.

3.6.6.2 Finite difference method

The most straightforward method for computing the gradient is to use finite differences, which consists of an approximation where the gradient is evaluated by varying each variable of the cost function. For instance, for a first-order finite difference scheme:

$$\nabla_j J(\mathbf{x}_0) \approx \frac{J(\mathbf{x}_0 + \epsilon \mathbf{e}_j) - J(\mathbf{x}_0)}{\epsilon}, \quad (3.86)$$

where $\mathbf{e}_j$ is the unit vector in the j -th direction, and ϵ is the differentiation step. ϵ should be sufficiently small to accurately estimate the gradient of the cost function. Usually, a first-order scheme is considered, since increasing the order increases the number of evaluations of the costly nonlinear model. The finite difference method is simple to implement, but becomes computationally expensive when the dimension of $\mathbf{x}_0$ is significant, since it requires one forward model evaluation per control variable. For high-dimensional systems, the gradient of the cost function can be estimated more efficiently with the adjoint method, when possible.

3.6.6.3 Adjoint method

To compute gradients more efficiently, one approach is to use an adjoint model, which can be derived using the TLM or Lagrange Multipliers [[Asch et al., 2016](#)]. An adjoint model propagates gradients backward in time, opposite to the forward model. It computes how changes in control variables affect the output (cost function). The adjoint method originates from optimal control theory [[Lions, 1971](#); [Le Dimet and Talagrand, 1986](#)], and allows the gradient with respect to control variables to be computed efficiently, regardless of the number of control variables. Calculating the gradient with the adjoint model is equivalent to finding the solution of the adjoint model. While the computational cost of solving the adjoint model is typically two to four times higher than that of the forward model, it requires the same order of magnitude of time to run it [[Griewank and Walther, 2008](#)]. Then, only two model evaluations are needed for computing the gradient: a forward model run and an adjoint model run.

Remark: While this approach drastically reduces the number of evaluations of the model for variational approaches, the adjoint of a high-dimensional system is difficult to construct. It is an intrusive method, where the derivation of the adjoint model should be updated any time there is a change in the direct model. The adjoint models are derived based on code derivation techniques [Giering and Kaminski, 1998], by hand [Weaver et al., 2003], or with automatic differentiation [Hascoet and Pascual, 2013]. The adjoint does not always exist for industrial codes because it is close to constructing and maintaining a second version of the model. In addition, automatic differentiation is not always feasible for large or poorly structured codes, making manual intervention and code refactoring necessary.

Note: The adjoint of the TELEMAC-2D flood model is not available, so the finite difference scheme is used to estimate the gradient cost function. In the parameter estimation problem where $\mathbf{x} = \mathcal{M}(\boldsymbol{\theta})$, the control variable is the input parameter vector $\boldsymbol{\theta}$. The dimension of $\boldsymbol{\theta}$ is lower than the dimension of the state. The dimension of the parameter in the 2D flood models used in this research varies from 2 to around a hundred. For costly model evaluation with a hundred parameters, the assimilation is complicated, but it remains possible with HPC.

3.7 Summary

This chapter introduces the mathematical tools used in this thesis to combine uncertain observations with an uncertain 2D flood model for calibrating floodplain roughness. In particular, we presented UQ, which is essential for studying the influence of floodplain roughness and its variability on the model's outputs. Techniques such as Sobol indices computed via PCE are used in the following chapters to identify which roughness parameters are influential in the floodplains.

Once the influential parameters are determined, DA techniques are applied to calibrate them. This chapter reviewed different approaches, from the KF to variational approaches. Based on this review, the 3D-Var technique is mainly used in the following. However, the EnKF and PF are also tested on a schematic case in Chapter 7 to study the influence of the choice of DA techniques on the roughness parameterization. Beyond the methods themselves, this chapter emphasizes the importance of carefully tuning the observation and background errors to accurately estimate the state or parameters of the numerical model.

The next chapters apply these tools in the context of flood modeling using satellite observations. Chapter 5 focuses on defining the appropriate quantity of interest when comparing observed and simulated flood maps. Then, the presented SA and DA frameworks are used to calibrate the floodplain roughness by assimilating flood maps in Chapter 6 using 3D-Var with the performance measures identified in Chapter 5. Finally, the DA techniques presented here are used in Chapter 7 to assimilate water depth fields.

Chapter 4

Flood extent and water depth from SAR observations

Résumé :

Ce chapitre vise à évaluer comment les différentes méthodes de prétraitement, de cartographie des inondations et d'estimation de la profondeur de l'eau affectent la précision et la fiabilité des données hydrauliques obtenues par satellite. L'étude évalue l'incertitude introduite à chaque étape du traitement des images SAR et vise à identifier des configurations robustes adaptées à une utilisation opérationnelle. Les méthodes de prétraitement (en particulier le filtrage du chatoiement), de cartographie des inondations et d'estimation de la profondeur de l'eau sont évaluées pour différentes configurations d'hyperparamètres.

Abstract:

This chapter aims to evaluate how various preprocessing, flood mapping, and water depth estimation methods affect the accuracy and reliability of satellite-derived hydraulic data. The study evaluates the uncertainty introduced at each stage of SAR image processing and identifies robust configurations suitable for operational use. The preprocessing (speckle filtering), flood mapping, and water depth estimation methods were all evaluated for varying hyperparameter settings.

Contents

4.1	Motivations	94
4.2	Preprint under revision	94
4.3	Additional Case Study	125
4.3.1	Methodology	125
4.3.2	Results	126
4.4	Graphical User Interface	130
4.5	Summary	131

4.1 Motivations

Chapter 2 introduced the principles of satellite remote sensing, and especially the possibility of identifying floods using SAR imagery. Section 2.2.3 detailed the principles of SAR and discussed its main sources of uncertainty. We also outlined the key steps involved in the use of SAR data for flood-related applications, including preprocessing of raw imagery, flood extent mapping, and water depth estimation. Chapter 3, which presented DA techniques, further emphasized the importance of quantifying observation uncertainties for assimilating data into numerical models.

Thus, this chapter focuses on extracting information from satellite imagery (especially from SAR) for calibrating and validating hydraulic models. The main contribution of this chapter is to evaluate various preprocessing, flood mapping, and water depth estimation methods with a particular focus on uncertainty evaluation. This chapter directly addresses the first research question introduced in Chapter 1:

Research Question 1

What information can be derived from satellite observations to calibrate 2D flood models, and how accurately can it be quantified?

Section 4.2 presents the proposed evaluation methodology for extracting flood extent and water depth from SAR imagery. This section is written in the form of a paper submitted to Natural Hazards and Earth System Sciences (NHESS). Then, Section 4.3 applies this evaluation methodology to a second case study on the Shannon River in Ireland. In this case, flood maps from the Copernicus EMS are used as a validation dataset in place of hydraulic simulations. Finally, Section 4.4 briefly introduces a user-friendly interface developed to support practitioners in extracting hydraulic information from satellite observations.

4.2 Preprint under revision

The following paper was submitted to Natural Hazards and Earth System Sciences (NHESS) on 31 July 2025 and was revised on 30 December 2025.

EVALUATING THE EFFECTS OF PREPROCESSING, METHOD SELECTION, AND HYPERPARAMETER TUNING ON SAR-BASED FLOOD MAPPING AND WATER DEPTH ESTIMATION

Jean-Paul Travert

EDF R&D, Laboratoire National d’Hydraulique et Environnement (LNHE), Chatou, France
Laboratoire d’Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France
jean-paul.travert@edf.fr

Cédric Goeury

EDF R&D, Laboratoire National d’Hydraulique et Environnement (LNHE), Chatou, France
Laboratoire d’Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France

Sébastien Boyaval

Laboratoire d’Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France
Inria, Paris, France

Vito Bacchi

EDF R&D, Laboratoire National d’Hydraulique et Environnement (LNHE), Chatou, France

Fabrice Zaoui

EDF R&D, Laboratoire National d’Hydraulique et Environnement (LNHE), Chatou, France

ABSTRACT

Flood mapping and water depth estimation from Synthetic Aperture Radar (SAR) imagery are crucial for calibrating and validating hydraulic models. This study uses SAR imagery to evaluate various preprocessing (especially speckle noise reduction), flood mapping, and water depth estimation methods. The impact of the choice of method at different steps and its hyperparameters is studied by considering an ensemble of preprocessed images, flood maps, and water depth fields.

The evaluation is conducted for two flood events on the Garonne River (France) in 2019 and 2021, using hydrodynamic simulations and in-situ observations as reference data. Results show that the speckle filtering method choice can significantly alter flood extent estimations with variations of several square kilometers. Additionally, the selection and tuning of flood mapping methods significantly affect performance. While supervised methods outperformed unsupervised ones, well-tuned unsupervised approaches (such as local thresholding or change detection) can achieve comparable results. The compounded uncertainty from preprocessing and flood mapping steps also introduces substantial variability in the water depth field estimates.

This study highlights the importance of considering the entire processing pipeline, encompassing preprocessing, flood mapping, and water depth estimation methods and their associated hyperparameters. Rather than relying on a single configuration, adopting an ensemble approach and accounting for methodological uncertainty should be privileged. For flood mapping, the method choice has the most influence. For water depth estimation, the most influential processing step was the flood map input resulting from the flood mapping step and the hyperparameters of the methods.

1 Introduction

Flood risk management largely benefits from accurate, timely, and spatially extensive observations of flood events. Satellite remote sensing allows the monitoring of large areas with increasing spatial and temporal resolution. Among the various types of satellite sensors, Synthetic Aperture Radar (SAR) sensors are particularly valuable for flood monitoring (Oberstadler et al., 1997; Bates, 2012). Unlike optical sensors, SAR sensors acquire images regardless of cloud coverage and daylight, making them suitable for detecting flood events with high spatial resolution (Tarpanelli and Benveniste, 2019). Flood mapping with SAR imagery exploits the interaction between the emitted signal and water surfaces, which typically results in dark spots in the images due to specular reflection where the smooth water surface mirrors the signal away from the sensor (Hostache et al., 2009; Mason et al., 2009; Martinis, 2010).

Numerous methods have been developed to extract flood maps from SAR imagery, including global and local thresholding methods (Chini et al., 2017; Mason et al., 2012), active contour models (Horritt, 1999), change detection methods (Giustarini et al., 2012; Bovolo and Bruzzone, 2007) or supervised classification approaches (Bentivoglio et al., 2022; Mateo-Garcia et al., 2021; Bonafilia et al., 2020). These flood maps provide essential data for model calibration and validation (Di Baldassarre et al., 2009; Montanari et al., 2009), and data assimilation in hydraulic forecasting systems (Hostache et al., 2010; Lai et al., 2014; Giustarini et al., 2011; Dasgupta et al., 2021).

Beyond generating flood maps, estimating water depth from SAR data is also valuable for hydraulic model calibration and validation (Hostache et al., 2009; Schumann et al., 2007; Betterle and Salamon, 2024). These estimations usually involve combining flood maps and ancillary datasets, such as Digital Elevation Models (DEMs) (Hostache et al., 2009; Schumann et al., 2007; Betterle and Salamon, 2024) or outputs from hydrodynamic simulations (Brown et al., 2016).

However, extracting hydraulic information such as flood maps and water depth fields from SAR images is subject to various sources of uncertainty. These include measurement noise (e.g., speckle), terrain-induced distortions, and vegetation and infrastructure influence. Furthermore, uncertainty arises from methodological choices in the processing workflow, such as noise filtering strategy, flood mapping methods, and hyperparameter settings. In this work, we refer to hyperparameters as user-defined configuration values that govern the behavior of a method (e.g., window size for speckle filtering, threshold level for flood classification) but are not learned from a calibration on observations such as model parameters in numerical flood models. Previous studies attempted to quantify these uncertainties. For example, Schumann et al. (2008) propagated geolocation uncertainties to derive ensembles of flood maps and associated water depth. Similarly, Martinis et al. (2015b) compared various operational flood mapping strategies to study their relevance in operational contexts. Landuyt et al. (2018) compared flood mapping methods with hyperparameter tuning across multiple flood events in Ireland and the UK (comparison to Copernicus Emergency Mapping Service flood maps), highlighting that method performance was highly variable depending on the study case and the fixed hyperparameters. Tupas et al. (2023), Chini et al. (2017) and (Ghosh et al., 2024) drawn the same conclusions when evaluating the role of parameterization in change detection, thresholding, and machine-learning based methods on the flood mapping outputs, respectively. However, flood studies rarely analyze the influence of preprocessing or hyperparameter tuning across the complete workflow, including flood mapping and water depth estimation. For instance, (Landuyt et al., 2018) assume a unique preprocessing strategy for the satellite images without evaluating its role in the subsequent analysis.

In this study, we proposed a comprehensive workflow for SAR-based flood analysis, evaluating the sensitivity of flood mapping and water depth estimation to different combinations of preprocessing, flood mapping, and water depth estimation methods. The objective is to quantify the uncertainty introduced at each stage of the SAR image processing and identify robust configurations for operational use. The preprocessing (speckle filtering), flood mapping, and water depth estimation methods were all evaluated for varying hyperparameter settings. This evaluation was carried out in an operational context with two flood events on the Garonne River in France in 2019 and 2021, with two Sentinel-1 SAR observations available for both flood events. All code and data are publicly available to facilitate reproducibility and further experimentation by the hydraulic modeling community at <https://github.com/jtravert/sar-flood-evaluation-framework>.

This article is structured as follows. Section 2 describes the methodology and study area, including the satellite data and validation datasets. Section 3 describes and applies the SAR image preprocessing steps, explicitly focusing on speckle filtering strategies. Section 4 reviews and evaluates flood mapping approaches. Section 5 presents the estimation of water depths from SAR-derived flood maps using DEMs. Section 6 discusses limitations of this study and their operational use for model calibration. Finally, Section 7 provides conclusions and implications for operational flood monitoring.

2 Methodology

The general workflow for processing SAR satellite images for flood applications is illustrated in Fig. 1.

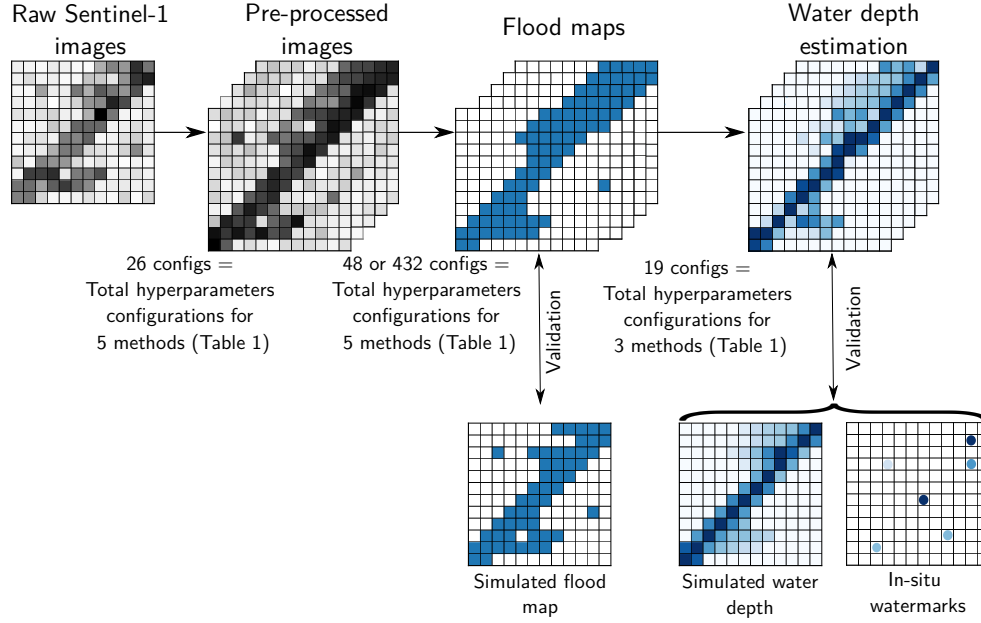


Figure 1: Workflow for processing Sentinel-1 (SAR) images to derive flood maps and water depth fields. Raw images are preprocessed using 26 hyperparameter configurations, then each image is turned into 48 flood maps (or 432 if using morphological operations). Each flood map is transformed into water depth fields using 19 configurations. The configurations are described in Table 1.

The workflow consists of the following steps:

1. **SAR image preprocessing:** Raw SAR images are first preprocessed to improve flood signal extraction. Five filtering methods were tested, including the Median filter, Lee filter (Lee, 1980), Lee Sigma filter (Lee et al., 2008), Frost filter (Frost et al., 1982), and the SAR2SAR deep-learning-based approach (Dalsasso et al., 2021). Each speckle filtering method was applied with a specific set of hyperparameters, resulting in 26 unique configurations: one configuration without any preprocessing (no hyperparameters), three configurations each for the Median and Lee filters, nine configurations each for the Frost and Lee Sigma filters, and one configuration for SAR2SAR.
2. **Flood mapping:** Each preprocessed image is used as input for flood mapping. Five methods were tested: global thresholding, local thresholding, active contour models, change detection, and supervised classification. These methods are also parameterized, resulting in an ensemble of flood maps for each preprocessed input. Post-processing morphological operations are optionally applied to remove isolated water pixels or small nonphysical holes caused by SAR speckle, geometric distortions, or flood map processing. In total, for each input preprocessed image, 48 flood mapping configurations are tested: one configuration each for the supervised classification models (no hyperparameters), two for global thresholding, two for change detection, six for active contour, and 36 for local thresholding. When using morphological post-processing, nine configurations were evaluated. With morphological operations, 432 flood mapping outputs (48×9) were generated per preprocessed image.
3. **Water depth estimation:** Each flood map generated in the previous step, together with a Digital Elevation Model (DEM) including the bathymetry for the river channel, is used to estimate water depth fields using three methods: Fw-DET (Cohen et al., 2019), FLEXTTH (Betterle and Salamon, 2024), and a cross-sectional hydraulic approach. For each input flood map, 19 water depth estimation configurations are tested: one configuration for the cross-section approach (no hyperparameters), nine for Fw-DET, and nine for FLEXTTH.

The preprocessing, flood mapping, and water depth estimation methods are described in more details in Sections 3, 4 and 5 respectively. All the methods and their respective hyperparameters are summarized in Table 1. Although validation datasets, such as hydrodynamic simulations, were used to support a better understanding of the methodology, the main focus of the study was to explore the variability in outputs resulting from these choices, rather than to perform a strict

validation. The range of the hyperparameters is based on classical values used in the literature to avoid non-physical hyperparameter values.

Category	Method/Filter	Hyperparameter	Hyperparameter values
Speckle filtering	No preprocessing	-	-
	Median	Window size	{3; 5; 7}
	Lee	Window size	{3; 5; 7}
	Lee Sigma	Window size	{3; 5; 7}
		Cumulative probability (ξ)	{0.7; 0.8; 0.9}
	Frost	Window size	{3; 5; 7}
		Damping factor (α)	{1; 2; 3}
	Deep-learning-based (SAR2SAR)	Model weights	Pre-trained model weights (fixed)
	Global thresholding	Threshold selection procedure	{Otsu, Kittler and Illingworth}
	Local thresholding	Minimum tile size in pixels	{100 × 100; 200 × 200}
Flood mapping		Ashman's D threshold	{1.9; 2.0; 2.1}
		Bhattacharyya coefficient	{0.98; 0.99}
		Surface ratio	{0.05; 0.1; 0.15}
	Active contour	Contour smoothness (α)	{0.05; 0.1; 0.2; 0.3; 0.4; 0.5}
	Change detection	Classification method	Global or local thresholding
	Supervised classification	Model	CNN or Random Forest (fixed weights)
	Morphological operations	Holes area in pixels	{10; 50; 100}
		Patches removal in pixels	{10; 50; 100}
	Fw-DET	Slope threshold	{no threshold; 5%; 10%}
		Number of smoothing iterations	{3; 5; 10}
Water depth estimation	FLEXTH	Slope threshold	{no threshold; 5%; 10%}
		Maximum number of neighbors	{5; 10; 20}
	Cross-section analysis	-	-

Table 1: Overview of speckle filtering, flood mapping, and water depth estimation methods used in this study, along with their associated hyperparameter sampling.

2.1 Study area and materials

The study area was the Garonne River between Tonneins and La Réole, in southwest of France (see Fig. 2). In this section of the river, the river width is around 250 m and the floodplain is 1-4 km wide, mainly comprises rural areas used for agricultural purposes. Since the end of the nineteenth century, the area has been equipped with dikes to protect urban areas and infrastructures. The floodplain, aside from the presence of dikes, displays minimal topographic variation.

2.1.1 Satellite observations

The study area was observed during two flood events in December 2019 and early February 2021 by the Sentinel-1 C-band Synthetic Aperture Radar (SAR) instrument at 5.405 GHz. The extent of the study area and the satellite image acquisitions are visualized in Fig. 2. Sentinel-1 “Ground Range Detected” products were downloaded from ASF Data Search Vertex (<https://search.asf.alaska.edu/>) in VH polarization. The role of polarization is beyond the scope of this study. Most flood mapping methods rely on the analysis of a single polarization image. Accordingly, the VH polarization was used since, in the literature, VH polarization better distinguishes flooded from dry areas than VV polarization (Henry et al., 2006). Two additional Sentinel-1 images, acquired under non-flooded conditions, were used as references for one of the flood mapping methods. For the 2019 event, the reference image was acquired on 10 December 2019. For the 2021 event, the reference observation was acquired on 28 January 2021. Each image consists of an $N_x \times N_y$ grid of pixels, with a spatial resolution of 10 × 10 m. The images are projected onto a common grid that covers the entire study area. Here, the image's dimensions are $N_x = 2644$ and $N_y = 2312$. All times reported below are given in Coordinated Universal Time (UTC).

2.1.2 Validation data

For comparing the observed flood maps and water depth fields derived from satellite images, observed water marks and stage gauging stations are available. These observations were also compared to simulated flood maps and water depth fields. The simulations are not the ground truth, but serve as a reference for comparison. The main goal of the study is to compare the variability of the outputs due to preprocessing, method choices, and hyperparameters, so the validation dataset is not the most important. We describe these datasets below.

Watermarks are visible traces left on buildings, trees, or other infrastructures during a flood event at the peak water level. For both the 2019 and 2021 flood events, watermarks were collected and made available on the French collaborative platform “Repères de Crues” (<https://www.reperesdecruces.developpement-durable.gouv.fr/>). For the

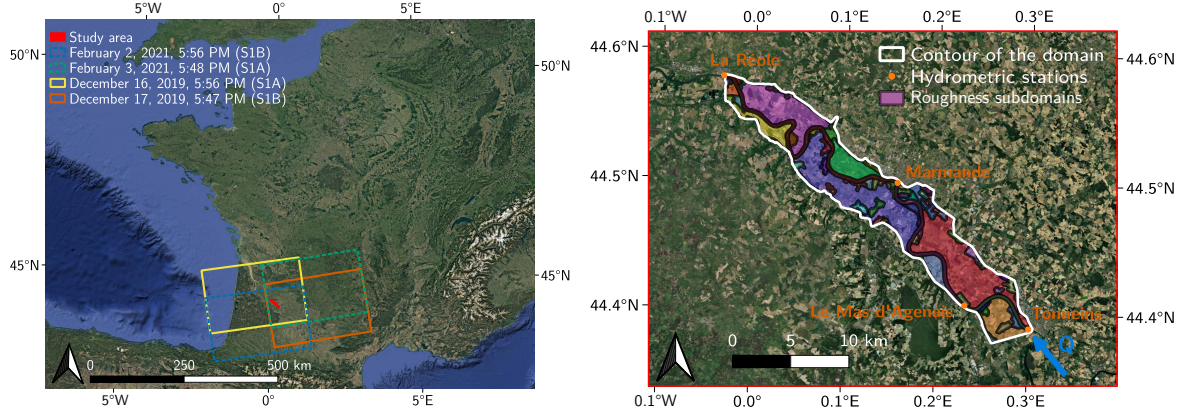


Figure 2: Visualization of the study area on the Garonne River in France with (a) the extent of the Sentinel-1 A (S1A) and Sentinel-1 B (S1B) acquisitions during 2019 and 2021 floods event and (b) a zoom-in on the study area with in-situ stage gauging stations and roughness subdomains.

2019 and 2021 flood events, 121 and 178 are available, respectively. For both events, satellite images were acquired near the flood peak (17 December 2019 and 3 February 2021), so the watermarks should coincide with the water depths extracted from the satellite images.

Three stage-gauging stations are available in the study area (Tonneins, Marmande, and La Réole). Discharge and water level data at these locations during the flood events are available on Vigicrues (<https://www.vigicrues.gouv.fr/>, a flood-monitoring service that collects watermarks during floods in France). The measured discharges at the three stations for both flood events are shown in Fig. 2.

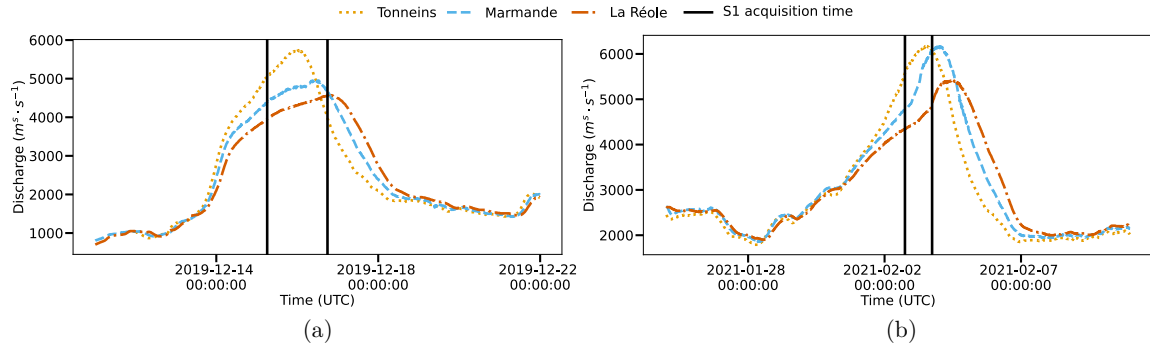


Figure 3: Measured discharge at gauging stations (Tonneins, Marmande, and La Réole) during the (a) December 2019 and (b) February 2021 flood event.

For flood simulations, a numerical solution of the Shallow Water Equations (SWEs) was used and solved with TELEMAC-2D, part of the openTELEMAC open-source hydrodynamic modeling system (www.opentelemac.org) (Hervouet, 2007). The SWEs are expressed in Cartesian coordinates where gravity acts uniformly in the vertical direction as $-g\vec{e}_z$:

$$\frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} = 0, \quad (1)$$

$$\frac{\partial(hu)}{\partial t} + \frac{\partial(hu^2)}{\partial x} + \frac{\partial(huv)}{\partial y} = -gh \frac{\partial \eta}{\partial x} + \nabla \cdot (h \nu_e \nabla u) - \frac{1}{\rho} \tau_{bx}, \quad (2)$$

$$\frac{\partial(hv)}{\partial t} + \frac{\partial(hv^2)}{\partial y} + \frac{\partial(huv)}{\partial x} = -gh \frac{\partial \eta}{\partial y} + \nabla \cdot (h \nu_e \nabla v) - \frac{1}{\rho} \tau_{by}. \quad (3)$$

The system unknowns are the water depth $h \geq 0$ and the depth-averaged velocity $u\vec{e}_x + v\vec{e}_y$ both functions defined as function of spatial coordinates x, y and time $t \in [0, T)$. The water surface is denoted by $\eta = h + z_b$, where $z_b(x, y)$ is the prescribed bottom elevation. A constant viscosity $\nu_e > 0$ is assumed, and the bed shear stress is expressed as $\tau_{bx}\vec{e}_x + \tau_{by}\vec{e}_y$ depending on the variables h, u , and v . The bed shear stress is computed using Manning-Strickler formulation (Manning et al., 1890):

$$\begin{cases} \tau_{bx} = \frac{\rho \cdot g \cdot u}{h^{1/3} \cdot K_s^2} \sqrt{u^2 + v^2} \\ \tau_{by} = \frac{\rho \cdot g \cdot v}{h^{1/3} \cdot K_s^2} \sqrt{u^2 + v^2} \end{cases}, \quad (4)$$

where K_s is the Strickler coefficient which varies spatially with x , and y (Morvan et al., 2008).

Upstream discharge was retrieved from the Tonneins stations for 11-21 December 2019 and 25 January-10 February 2021 and reported in Fig. 3. A rating curve was imposed at the downstream boundary. The simulations were initialized at $t = 0$ with a base flow of $800 \text{ m}^3/\text{s}$ for the 2019 event, and $2300 \text{ m}^3/\text{s}$ for the 2021 event. One simulation was conducted for each flood event. The Strickler values were calibrated in previous studies for the river channel (Besnard and Goutal, 2011; El Garroussi et al., 2019; Nguyen et al., 2022), and are fixed in the floodplains based on the land use (Chow et al., 1988). The spatial distribution of Strickler values used in the simulation is reported in Fig. 2 and their values are described in Table 2 taking the median value of the Strickler value ranges of Design of Experiment 1. Bathymetry in the channel was reconstructed from 70 cross-sectional elevation profiles, and floodplain topography was obtained using data from IGN (the French National Institute of Geographic and Forest Information) combined with aerial imagery. The reader can refer to Besnard and Goutal (2011) and Traver et al. (2025) for more information on the construction and parameterization of the numerical model.

3 Satellite images preprocessing

Synthetic Aperture Radar (SAR) images, such as those acquired by Sentinel-1, require several preprocessing steps to correct for sensor geometry, normalize radiometric responses, and suppress speckle noise. The processing chain adopted in the present study is presented in Fig. 4 and comprises sequential operations: application of orbit files, removal of thermal and border noise, radiometric calibration, speckle filtering, and terrain correction using the Range-Doppler method. In the processing workflow for terrain correction and georeferencing, a Digital Elevation Model of the study area at a 1 m resolution is used. Two alternative workflows were used depending on the speckle filtering strategy since the deep-learning-based speckle filtering (SAR2SAR method) was trained on images without preprocessing, while the other traditional filters were applied on calibrated images.

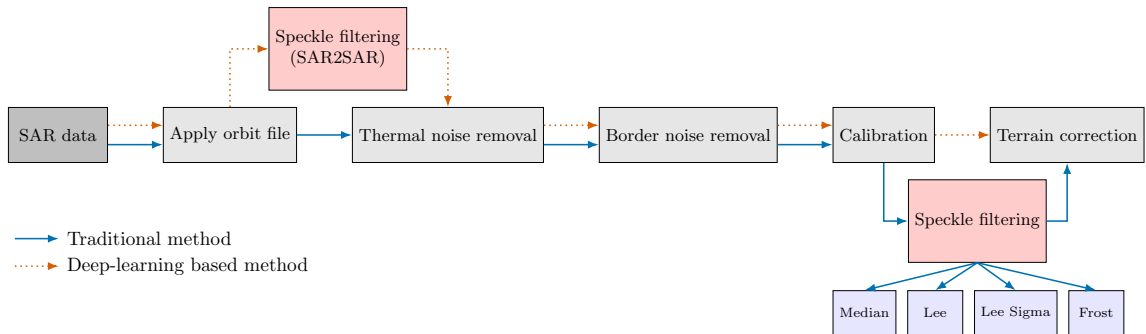


Figure 4: Preprocessing workflow for Sentinel-1 SAR images.

In SAR imagery, speckle noise which arises from the coherent summation of scattered electromagnetic waves can be present. It resembles salt and pepper with dark and bright pixels (see Fig. 5). It alters the statistical properties of the image, preventing them from maintaining a consistent mean radiometric level in homogeneous areas (Bruniquel and Lopes, 1997). Speckle noise is modeled with a randomly fluctuating variable, such as (Goodman, 1976):

$$I_{i,j} = R_{i,j} \cdot S_{i,j}, \quad (5)$$

where $I \in \mathbb{R}^{N_x \times N_y}$ denotes the observed intensity (raw image), $R \in \mathbb{R}^{N_x \times N_y}$ the true radar backscatter, $S \in \mathbb{R}^{N_x \times N_y}$ the speckle component, and i, j the pixel locations. $S_{i,j}$ and $R_{i,j}$ are assumed to be statistically independent.

3.1 Speckle filters methods

To mitigate the effects of noise, several speckle filters that aim to recover R the true radar backscatter while preserving image structures have been proposed in the literature. A review on speckle filtering methods is available in Deledalle et al. (2014) and Lee and Pottier (2017). In this study, we selected a representative set of widely used traditional filters such as Frost, Lee, Lee Sigma, and Median filter, due to their simplicity and strong track record in SAR image denoising. Additionally, we included a recent deep learning-based method to evaluate potential performance improvements from modern data-driven approaches. The characteristics of these filters are summarized in Table 1. In the following sections, each method is described in more detail.

The deep-learning based model (SAR2SAR model) was implemented via the deepdespeckling Python library available at <https://pypi.org/project/deepdespeckling/> (Dalsasso et al., 2021). The other traditional filtering operations were conducted using the ESA SNAP toolbox. The complete preprocessing chain was automated using SNAP's Graph Processing Tool (GPT) and Python scripting, facilitating large-scale processing of SAR image stacks.

3.1.1 Median filter

The Median filter replaces the center pixel with the median value of all pixels within a local scanning window, such that:

$$\hat{R}_{i,j} = \text{median} \{I_{m,n} \mid i - k \leq m \leq i + k, j - k \leq n \leq j + k\}, \quad (6)$$

with a window size of $2k + 1$ with k a non-negative integer. In this study, the window size is treated as a hyperparameter. The median filter is effective at removing isolated spot noise, but it tends to blur edges and erase thin linear features (Lee, 1983). To address this limitation, other approaches referred to as adaptive filters have been developed that incorporate local image statistics to better preserve structural details.

3.1.2 Lee filter

The Lee filter (Lee, 1980) accounts for the local statistics of the image within a moving window. From the multiplicative noise definition in Equation 5, the mean and variance of R can be estimated, such as:

$$\bar{R}_{i,j} = \bar{I}_{i,j}, \quad \text{Var}(R_{i,j}) = \frac{\text{Var}(I_{i,j}) - \bar{I}_{i,j}^2 \text{Var}(S_{i,j})}{\text{Var}(S_{i,j}) + \bar{S}_{i,j}^2}, \quad (7)$$

where the quantities $\bar{I}_{i,j}$ and $\text{Var}(I_{i,j})$ are the local mean and variance of pixel intensities within a window of size $2k + 1$ centered at (i, j) . In this study, the window size is treated as a hyperparameter. The Lee Filter assumes a linear estimator and using the local mean and variance within each scanning window, the estimator is written as:

$$\hat{R}_{i,j} = \bar{R}_{i,j} + \frac{\text{Var}(R_{i,j})}{\text{Var}(I_{i,j})} (I_{i,j} - \bar{S}_{i,j} \bar{R}_{i,j}). \quad (8)$$

The Lee filter aims to reduce speckle in homogeneous areas while preserving image details, such as edges and fine structures, in areas of high variance.

3.1.3 Lee Sigma filter

The Lee filter effectively reduces speckle noise but can cause blurry edges and loss of detail in heterogeneous areas. To address this issue, the Lee Sigma filter (Lee, 1983) was introduced. It calculates local statistics similarly to the Lee filter but using only the pixels whose intensities fall within a range to exclude outliers. The range is defined as $[\bar{R}_{i,j} - 2\sqrt{\text{Var}(S_{i,j})}\bar{R}_{i,j}, \bar{R}_{i,j} + 2\sqrt{\text{Var}(S_{i,j})}\bar{R}_{i,j}]$. Because the a priori mean, $\bar{R}_{i,j}$, is unknown, it is approximated by $\bar{I}_{i,j}$, the value of the center pixel.

An improved version (Lee et al., 2008) relaxes the usual range and introduces a new interval (I_1, I_2) that meets two conditions:

- The interval captures a fixed cumulative probability ξ :

$$\xi = \int_{I_1}^{I_2} p(I) dI, \quad (9)$$

where $p(I)$ is the empirical probability distribution of the pixel intensity computed on the scanning window.

- The mean within the interval must match the overall mean:

$$\bar{I} = \frac{1}{\xi} \int_{I_1}^{I_2} I p(I) dI. \quad (10)$$

A value of $\xi = 0.8$ or 0.9 is often used, but a lower value of ξ may be selected to preserve SAR image texture information and prevent potentially over-smoothing details (Lee et al., 2008). Two hyperparameters are used for the Lee Sigma filter, the window size and ξ .

3.1.4 Frost filter

Similarly to the Lee filter, the Frost filter (Frost et al., 1982) is based on the local statistics of the images and the multiplicative noise model. The Frost filter replaces a pixel value with a weighted sum of the values of its neighbors within a moving scanning window, such that:

$$\hat{R}_{i,j} = \sum_{m=-k}^k \sum_{n=-k}^k w_{m,n} \cdot I_{i+m,j+n}, \quad (11)$$

where the weights $w_{m,n}$ decrease with distance from the pixel of interest with an exponential decay controlled by a damping factor α , such that:

$$w_{m,n} = \frac{e^{-\alpha d(m,n)}}{\sum_{m=-k}^k \sum_{n=-k}^k e^{-\alpha d(m,n)}}, \quad (12)$$

with $d(m,n)$ the Euclidean distance of pixels at position m,n from the pixel at the center of the scanning window. For a pixel center located at i,j the distance is defined as $d(m,n) = \sqrt{(m-i)^2 + (n-j)^2}$. For the Frost filter, two hyperparameters are considered in this study, the window size and the damping factor α .

3.1.5 Deep-learning based filtering

Due to the difficulty of removing speckle noise with empirical models based on image statistics, deep learning approaches are increasingly used. SAR2SAR (Dalsasso et al., 2021) is a deep learning-based despeckling method using a U-Net architecture (Ronneberger et al., 2015) which aims to predict speckle noise. The U-Net is trained with synthetic speckle realizations, then fine-tuned on real SAR image pairs from a time series. After training, weights are used to compute the denoised image as a function of input pixel intensities. In this study, we used a pre-trained model with weights available at <https://gitlab.telecom-paris.fr/ring/sar2sar>. For SAR2SAR, no hyperparameters were tested.

3.2 Application of speckle filtering

We applied the five filtering methods for all hyperparameter configurations to the four Sentinel-1 SAR images (and to the reference non-flooded images). In our implementation, the computation time for the speckle filtering step ranged from a few seconds for the Lee and Median filters, to approximately 20-30 seconds for Lee Sigma and Frost filters, and up to around 4 min for SAR2SAR method, on Intel(R) Core(TM) i7-11850H @ 2.50GHz processor. Visual inspection and quantitative evaluations were carried out to compare the outputs of each method and their variability due to hyperparameter settings. Figure 5 illustrates how these filters (for one hyperparameter configuration) influence the backscatter for the same Sentinel-1 image.

The raw Sentinel-1 image (see Fig. 5a) exhibited significant speckle noise, appearing as salt and pepper granular regions. All tested filter configurations reduced the noise to varying extents, with variable success for preserving the edges. The Median and Lee filter (see Fig. 5b-c) reduced speckle noise but significantly blurred structural details along curvilinear structures. The Lee Sigma and Frost filters (see Fig. 5d-e) better balance noise reduction and detail preservation, effectively preserving both edges and textures. The SAR2SAR approach (see Fig. 5f) reduced the speckle noise while visually preserving edge features.

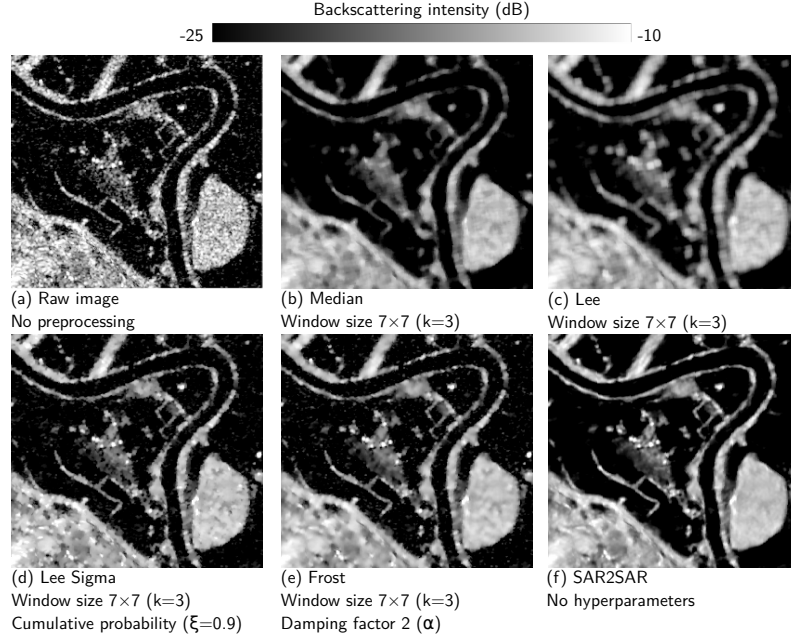


Figure 5: Backscatter intensity results from applying various speckle filters on a Sentinel-1 image in VH polarization.

The amount of speckle reduction can be quantified by calculating the Equivalent Number of Looks (ENL) over a quasi-homogeneous area, defined as:

$$\text{ENL} = \frac{\bar{\hat{R}}^2}{\text{Var}(\hat{R})}, \quad (13)$$

where $\bar{\hat{R}}$ is the mean estimated intensity over a homogeneous area, and $\text{Var}(\hat{R})$ is the variance of the estimated intensity over that same area. ENL estimates the signal-to-noise ratio, and the higher the ENL, the better the speckle suppression.

The ENL was computed for two zones, including a dry vegetated region and flooded region in blue and orange, respectively, in Fig. 6, for all the preprocessed images (corresponding to 26 configurations in total according to Table 1 plus one configuration without filtering) and for the four satellite images. The results are reported in Fig. 7, where each point corresponds to the ratio of the ENL between one of the preprocessed images over the non processed image. All filtering methods contributed to speckle reduction, as indicated by the systematically higher ENL ratio. It can be noted that the reduction in speckle noise differs between the dry and flooded areas, with better improvements in the dry region. The SAR2SAR achieved the highest ENL values across the different filtering methods (see Fig. 7a). Traditional statistical filters, such as the Lee and Frost filters, also significantly improved ENL for some hyperparameter configurations (largest window size and higher damping factor). The high variability of the ENL for the Lee, Median, and Frost filters underlined the significant impact of the method's hyperparameters on speckle reduction.

The quantitative analysis of edge and structure preservation is difficult because fine structures are close to the speckle noise spatial resolution (Lee et al., 1994). Edge preservation was not quantitatively evaluated in this study. The SAR2SAR approach gave the best quantitative results on noise reduction in homogeneous areas while visually preserving edges and structures. Traditional filtering methods had similar ENL values for some hyperparameter settings, showing their potential to reduce speckle noise but at the cost of preserving details (edge blurring) and the need for hyperparameter tuning.

Speckle reduction aims to improve the interpretability of the images and the extraction of information. Thus, the role of speckle reduction is analyzed in more detail in Section 4 and 5 to show its impact on interpreting SAR images and how the variability in speckle reduction changes the output.

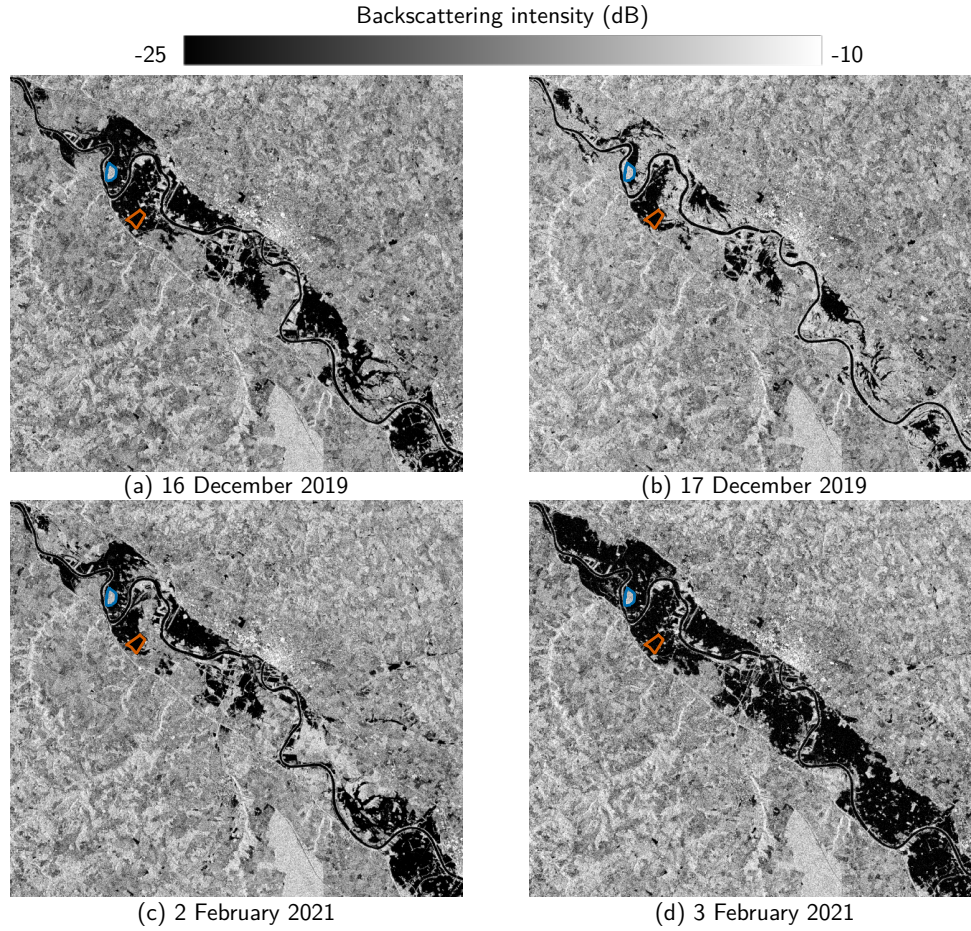


Figure 6: Four raw Sentinel-1 images in VH polarization of the 2019 and 2021 flood events on the Garonne River in France. The blue and orange polygons are used for the Equivalent Number of Looks computation.

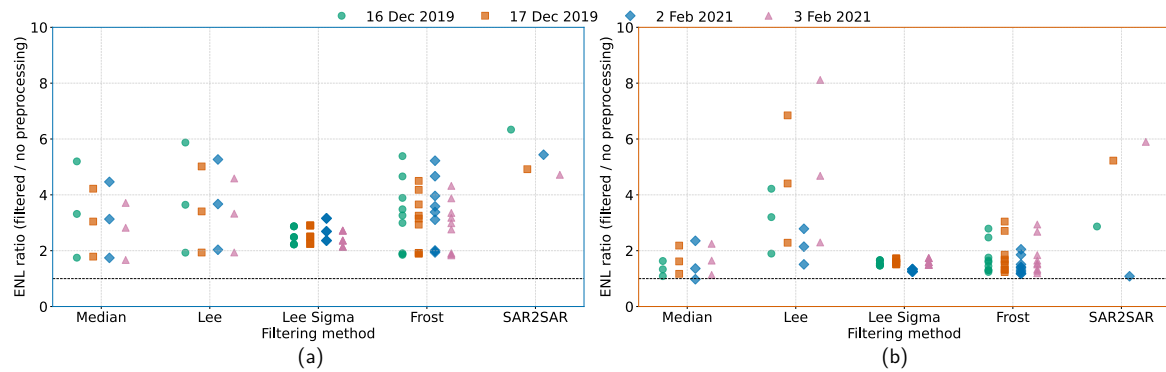


Figure 7: Equivalent Number of Looks (ENL) ratio (between images with and without preprocessing) for the four satellite images with different preprocessing for the dry homogeneous area (a) and flooded homogeneous area (b).

4 Flood mapping

In this study, we generated an ensemble of flood maps by applying several flood mapping methods across a range of hyperparameters, listed in Table 1. The methodology for flood map generation is illustrated in Fig. 8, and follows this general workflow:

1. **Input:** Preprocessed SAR images.
2. **Method selection:** Global thresholding, local thresholding, active contour models, change detection, or supervised classification (Convolution Neural Networks, Random Forest). Each method is evaluated over a range of hyperparameters (see Table 1). For instance, two methods for evaluating the threshold parameter are tested for global thresholding.
3. **Flood map generation:** Each parameter configuration results in a flood map $O \in \{0, 1\}^{N_x \times N_y}$.
4. **Post-processing (optional):** Morphological operations are applied on the flood maps to fill small holes and remove small elements not connected to the flood.

4.1 Methods

4.1.1 Global thresholding

Global thresholding of SAR images is a straightforward method to separate foreground (e.g., flooded pixels) from background (e.g., dry pixels) widely used in flood mapping for their simplicity (Martinis et al., 2009; Schumann et al., 2010; Pulvirenti et al., 2013). This approach assumes that the two-pixel classes (e.g., flooded and dry pixels) can be separated with a threshold value for the whole image. The threshold value is usually determined automatically with various methods (Sezgin and Sankur, 2004).

In this study, we evaluated the impact on the flood mapping of two widely used global thresholding methods: Otsu’s method (Otsu, 1979) and the Kittler and Illingworth (KI) method (Kittler and Illingworth, 1986). Complementary details on both methods are provided in Appendix A.1.

4.1.2 Local thresholding

Global thresholding assumes that the histogram is bimodal, which may not hold when the number of flooded pixels is low or due to spatial variability in backscatter caused by terrain or surface conditions across the image. To address this limitation, local-based thresholding was introduced (Martinis et al., 2015a; Twele et al., 2016; Chini et al., 2017). Images are divided into smaller tiles to identify bimodal histograms representing flooded and dry classes. From these identified sub-tiles, a threshold is computed with global thresholding. Following Chini et al. (2017), we used quad-tree decomposition to recursively split the image until reaching a minimum tile size (to guarantee statistical representativeness). A tile is eligible for thresholding if its histogram is bimodal, the distributions are normally distributed, and both classes are sufficiently represented. These conditions are evaluated by computing three qualitative coefficients (Ashman D, surface ratio, Bhattacharyya) described in Appendix A.2.

This study evaluated the impact on the flood mapping of the minimum tile size and of the three coefficients for different values reported in Table 1.

4.1.3 Active contour models

A classical active contour model is based on the Chan-Vese segmentation method (Chan and Vese, 2001), which is inspired by the Mumford–Shah model (Mumford and Shah, 1989). This approach determines a contour C that segments the image into two regions so that the pixel intensities are approximately constant within each region. The segmentation problem is formulated as the following minimization:

$$\min_C \alpha \text{Length}(C) + \nu \text{Area}(\text{inside}(C)) + \lambda_1 \sum_{(i,j) \in \text{inside}(C)} (\hat{R}_{i,j} - c_1)^2 + \lambda_2 \sum_{(i,j) \in \text{outside}(C)} (\hat{R}_{i,j} - c_2)^2, \quad (14)$$

where $\alpha \geq 0$ controls the smoothness of the contour, $\nu \geq 0$ penalizes the area inside the contour, $\lambda_1, \lambda_2 > 0$ are weighting the data fidelity inside and outside the contour, respectively, and c_1 and c_2 are the average intensities inside and outside the contour C , respectively. The algorithm is adapted so that $c_1 < c_2$, forcing flooded pixels with lower backscattering to be inside C .

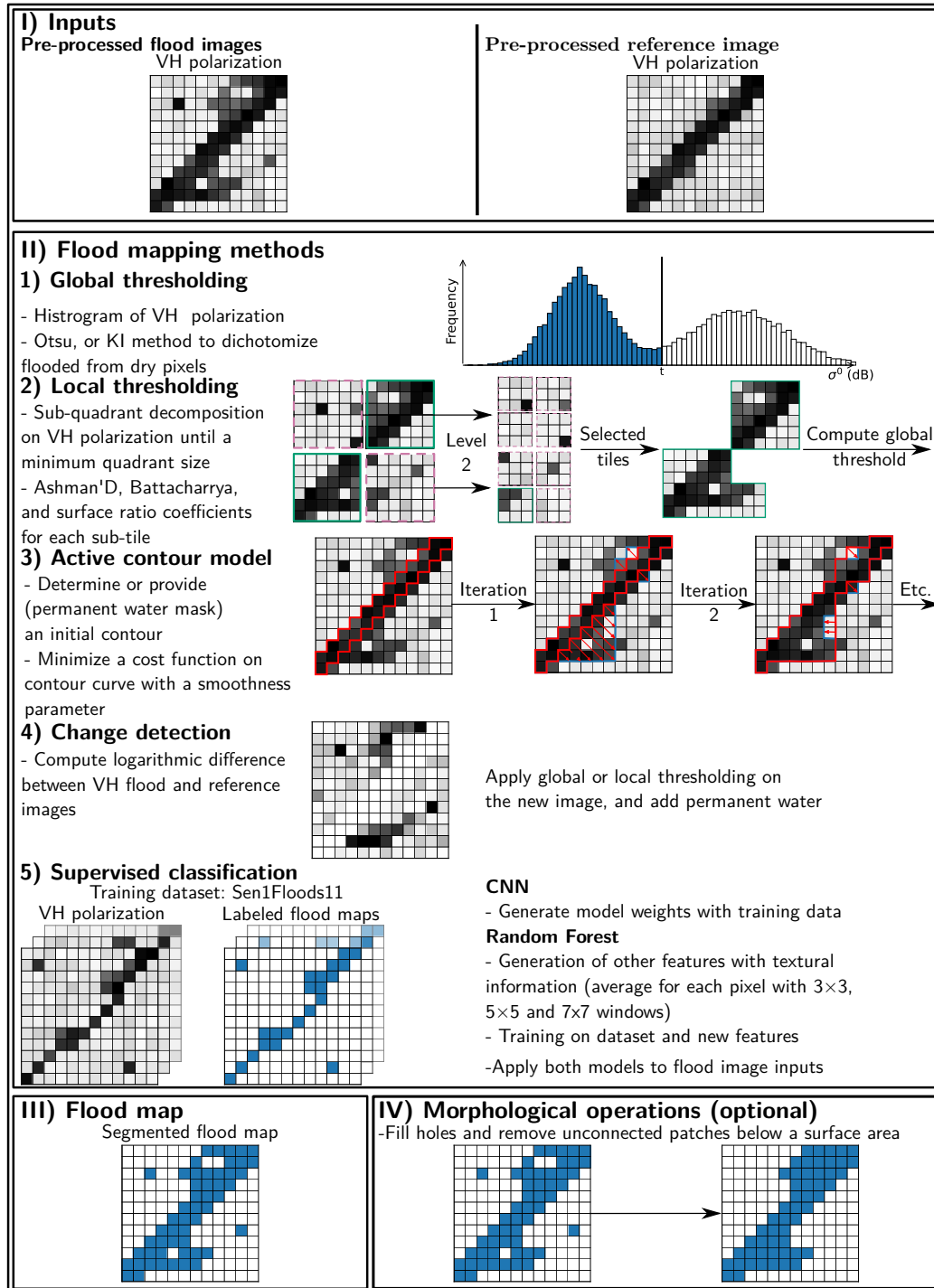


Figure 8: Schematic overview of the flood mapping methodology.

In this study, we considered α as a hyperparameter, while the remaining parameters were fixed to $\lambda_1 = 2$, $\lambda_2 = 1$, and $\nu = 0$ for computational cost reasons. We chose to constrain the problem with $\lambda_1 > \lambda_2$ to favor the identification of regions with lower mean intensities (typically associated with flooded areas in SAR imagery) as the foreground while assigning higher-intensity regions to the background. Active contour models require an initial contour as a starting condition for the segmentation process. Horritt (1999) suggests selecting the initial contour manually. In this study, the contour was initialized using a map of permanent water bodies from the Global Flood Monitoring Service with a 10 m resolution (<https://global-flood.emergency.copernicus.eu/>) (Martinis et al., 2022).

4.1.4 Change detection

Change detection methods require at least two satellite images of the same area, one during the flood and the others before or after the event. The output of change detection methods depends on the reference image without floods, so the reference image should be chosen carefully (Hostache et al., 2012). Although there are methods using a stack of images (Clement et al., 2018), we focused on a pair of images. First, we computed the logarithmic difference between the two images because of the multiplicative character of speckle noise (see Equation 5) (Bazi et al., 2005). Flooded pixels are then classified using global thresholding described in Section 4.1.1 with consistent parameter sampling. Similarly to global thresholding, the hyperparameter in the change detection method is the method to find the threshold on the image (Otsu or KI).

4.1.5 Supervised classification

Two supervised classification methods, widely used for flood mapping based on labeled training data (Zhao et al., 2020; Bentivoglio et al., 2022) are used in this work: Random Forests (RF) (Breiman, 2001) and Convolutional Neural Networks (CNN) (LeCun et al., 2015). For the RF approach, the pixel intensity is used as a feature. With only one band (VH polarization), the prediction capability of RF is limited. We added textural information by providing the mean value of each pixel with window sizes of 3, 5, and 7 pixels, resulting in four features.

Both models are trained with the Sen1Floods11 dataset (Bonafilia et al., 2020) (including 11 flood events with various geographic conditions) using VH polarization band and hand-labeled data dichotomizing flooded from dry regions. The trained model weights were used to generate flood maps from satellite imagery based on pixels' backscattering intensity. For reproducibility, the weights of both models are available at: https://github.com/jtravert/sar-flood-evaluation-framework/tree/main/sources/1_FloodExtent/methods/MLweights. The training procedure for the CNN was adapted from Bonafilia et al. (2020). The Random Forest classifier was trained for various numbers of estimators, maximum depth, and minimum leaf number with Bayesian optimization using the Optuna Python library (Akiba et al., 2019) (<https://pypi.org/project/optuna/>). For supervised classification methods, no hyperparameters are considered, as their hyperparameters were optimized during the training procedure using Bayesian inference.

4.2 Post-processing of flood maps

After applying flood mapping methods, the flood maps can still contain artifacts caused by measurement noise, speckle, or processing errors. To improve the spatial coherence of flood maps, morphological operations (standard methods used in image processing) can be used to refine the flood maps. These operations include hole filling, which fills small gaps within flooded regions, and removal of isolated patches not physically connected to the flooded areas.

In this study, the influence of these operations was studied for holes and small patches of water from $1,000 \text{ m}^2$ to $10,000 \text{ m}^2$. We applied morphological operations by filling holes of less than 10, 50, or 100 pixels and removing unconnected patches of less than 10, 50, or 100 pixels, representing nine configurations. The number of filled/removed pixels is the hyperparameter for morphological operations.

4.3 Results

Flood maps were generated using the five flood mapping methods with the hyperparameter settings detailed in Table 1. For each satellite acquisition, 1,248 generated flood maps were produced (26 speckle filtering configurations $\times$ 48 flood mapping setups). These flood maps were optionally processed with morphological operations (nine configurations), resulting in 10,998 additional possible flood map variations for each satellite acquisition. In our implementation, the computation time for mapping a flood on a single preprocessed image was a few seconds for all methods, except for active contour, which required 5 to 10 minutes (depending on the smoothing hyperparameter), on Intel(R) Core(TM) i7-11850H @ 2.50GHz processor. Each generated flood map was compared against hydrodynamic simulations, as described in Section 2.1.2. The evaluation process comparing observations to simulations focused on floodplain areas,

excluding permanent water bodies and exclusion zones. The exclusion layer indicates areas where the satellite sensor cannot accurately discriminate flooded areas from dry areas due to water-look-alike conditions (e.g., roads, dikes, wetlands). The exclusion mask on the Garonne domain was retrieved from the Global Flood Monitoring service and the generation process is described in Wagner et al. (2026).

At the grid points, a hit occurs when a pixel is correctly predicted to be flooded (True Positive (TP)), and a correct rejection (True Negative (TN)) occurs when it is correctly predicted to be dry. A false alarm (False Positive (FP)) and a miss (False Negative (FN)) occur when a pixel is simulated as flooded or dry, respectively, whereas the opposite is observed. Two standard metrics for comparing flood maps (Hunter, 2005; Grimaldi et al., 2016) were used:

- **Accuracy:** Proportion of correctly predicted pixels (both flooded and non-flooded) relative to the total number of pixels:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (15)$$

- **F1-score:** The F1-score is defined as:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \quad (16)$$

Both metrics range from 0 to 1, with higher values indicating better performance. The flooded area in the evaluation zone is also computed to evaluate the variability of the flood map outputs.

4.3.1 Flood mapping methods evaluation

First, the images generated without applying morphological post-processing operations are analyzed. Figure 9 shows the variability of the Accuracy, F1-score and flooded area metrics across the four satellite acquisitions. Each box plot represents the range of metric variations of a flood mapping method due to varying input preprocessing and hyperparameter settings. For each method, box plots are displayed in different colors for the four satellite images. These box plots should be compared with one another for the same acquisition date to assess method performance variability under consistent conditions.

CNN and Random Forest methods (supervised classification) outperformed unsupervised methods on median F1-score and Accuracy. The CNN approach presented the highest median performance (between 0.80 and 0.94 for F1-score) with a narrow inter-quartile range (<0.005) in three out of the four acquisition dates (except for 17 December 2019 acquisition). The Random Forest also performed well (F1-score median between 0.76 and 0.93), but with slightly lower scores and higher variability. Both supervised classification methods exhibit low variance because of the absence of hyperparameter tuning, leaving speckle filtering as the main source of variation. In contrast, active contour models exhibited larger variability in Accuracy and F1-score across all dates. It underlined the sensitivity to either the smoothing parameter (α) or image preprocessing. While they occasionally achieved reasonable performance for specific configurations, their inconsistency (e.g., F1-score inter quartile range between 0.27 and 0.38 on the 16 December 2019) makes them less reliable in operational settings. This method can work only after careful manual tuning, limiting its operational appeal. Active contour models are usually better for local scales and simple flood patterns around the initial contour. Thresholding methods (global and local) showed moderate performance and low variability. Additional configurations with a larger hyperparameter space showed that the thresholding methods were not very sensitive to their hyperparameters. For some configurations, local thresholding resulted in comparable metric values to supervised classification while being easy to use. Finally, change detection methods provided similar moderate results regarding F1-scores and Accuracy. Their variance was relatively low but was higher for some acquisition dates (17 December 2019 and slightly 3 February 2021), likely due to differences in magnitude change between the pre- and during-flood images, with smaller flood extent on 17 December 2019. The supervised classification method provided the best trade-off between accuracy and robustness. When training data or model weights are unavailable, local thresholding or change detection methods remain attractive because they require almost no user input and result in similar scores with appropriate tuning.

Figure 9 also displays the flooded area for all flood mapping methods and satellite acquisitions. In the evaluation zone, containing the simulating domain without the exclusion zones and the minor bed representing a total of 100,2 km², the simulated flooded areas were 78.78 km² (16 December 2019), 52.32 km² (17 December 2019), 57.70 km² (2 February 2021) and 86.41 km² (3 February 2021). CNN produced the largest flooded area estimates, with the smaller discrepancy to the simulated flooded area. In contrast, the other methods highly underestimated the flooded areas systematically relative to the simulations. In general, outputs from hydrodynamic models provide smooth flooded surface with hydraulic connectivity. On the contrary, for the generated flood maps from SAR imagery, floods in vegetated or urban areas or between the dikes are often not well detected which can cause an underestimation of the

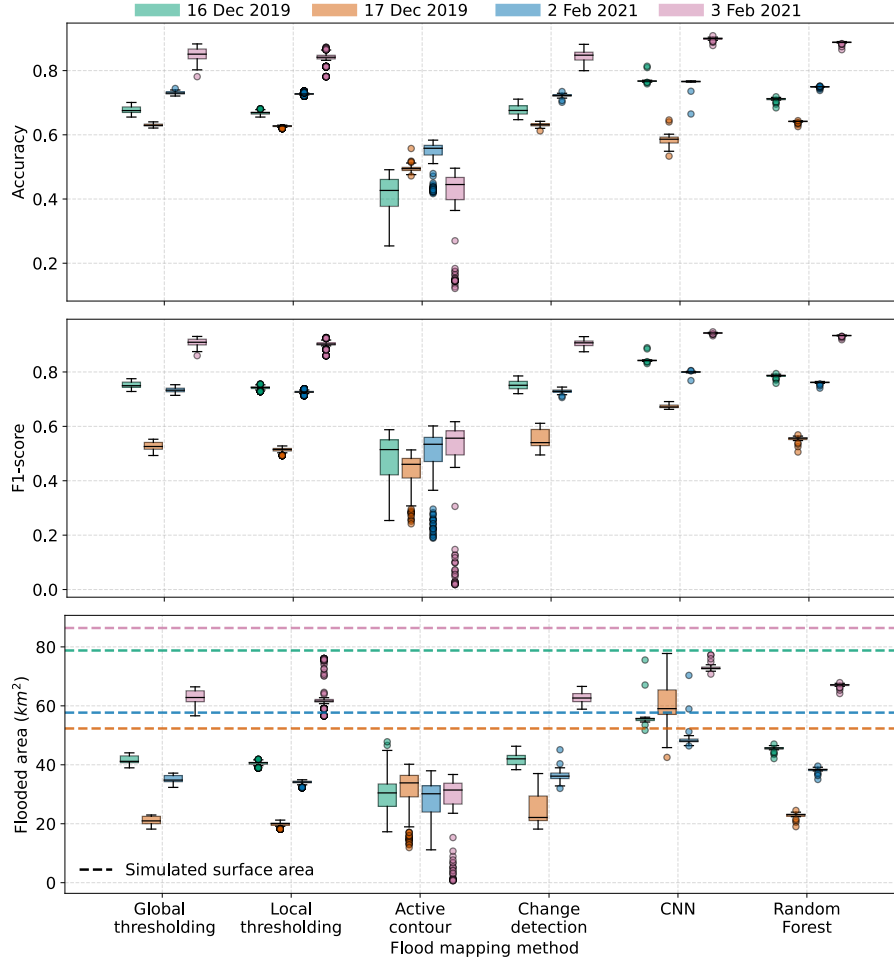


Figure 9: Accuracy, F1-score and flooded area for the generated flood maps for four satellite images and six flood mapping methods compared against their respective simulated flood maps.

flooded area, and the flood extent is more segmented and less smooth. On the one hand the simulations are conservative and overestimate the flooded area while the extracted flood maps underestimate the flooded areas. However, the main goal should be to match the geometric pattern and have a similar flood map geometry between the simulations and the observations even if there are small holes in the observed flooded area. The estimated flooded area can differ by 20 to 30 km² between methods, highlighting significant differences in flood map extraction approaches. Additionally, for a given method, the impact of hyperparameters or input images can be significant, with variations of 5 to 10 km² for active contour or change detection and CNN for 17 December 2019.

4.3.2 Impact of preprocessing

In most flood studies, the preprocessing of speckle noise is often assumed to be deterministic, with only one method (usually the Lee or Lee Sigma filter) (Di Baldassarre et al., 2009; Landuyt et al., 2018); however, the influence of speckle filtering choice on flood mapping can be significant. The preprocessing strategies should be evaluated for fixed configurations (flood mapping method and hyperparameter) to study the impact of preprocessing alone. For instance, the effect of preprocessing for CNN-generated flood maps is reported in Fig. 10. It highlighted the impact of preprocessing with a highly variable flooded area for Median filtering, depending on the window size of the filter (the only parameter for the Median filter). With filtering, the flooded area is larger than that of images without preprocessing. Between the different preprocessing methods, the flooded area has important variations showing the impact of the preprocessing method on the generated flood maps.

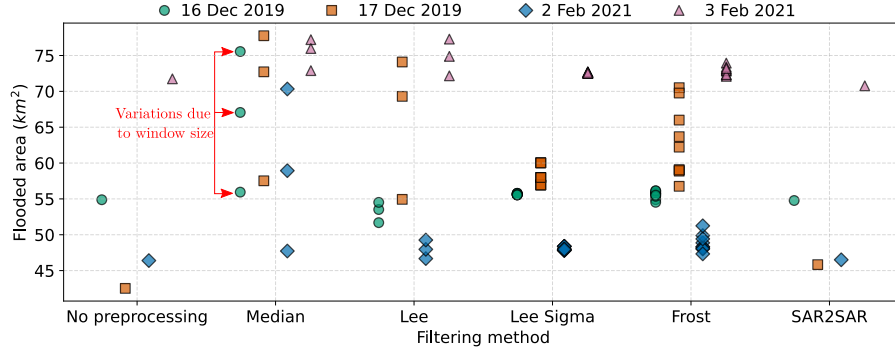


Figure 10: Impact of different filtering methods on the flooded area for flood maps generated using Convolutional Neural Networks.

Figure 10 illustrates the impact on the flooded area of the different filtering methods with a single flood mapping configuration. In total, Table 1 lists 48 configurations (2 for global thresholding and change detection, 36 for local thresholding, 6 for active contour, and one for each supervised classification method). For the Median, Lee, Lee Sigma, and Frost filters, we analyzed the impacts of their hyperparameters on these 48 configurations. The difference between the minimum and maximum value for each metric was calculated for every filter configuration (e.g., for the CNN configuration on 2 February 2021, the range of variation between the min and max values for the Median filter was 23 km²). Figure 11 presents the min-max variation in F1-score and flooded area due to speckle filtering hyperparameters for 2 February 2021. Each flood mapping configuration is visualized with a point with distinct markers for the different methods. The results indicate that the sensitivity to speckle filter hyperparameters varies by flood mapping method. For instance, the active contour method exhibits high sensitivity to the Median filter's window size, with variations ranging from 4 to 16 km². In contrast, global and local thresholding methods exhibit minimal sensitivity, with variations of less than 1 km². Overall, the Lee Sigma filter exhibited the lowest variability across all configurations. Local thresholding, change detection, and Random Forest configurations are more sensitive to Lee filter or Frost filter hyperparameters than Lee Sigma or Median filters.

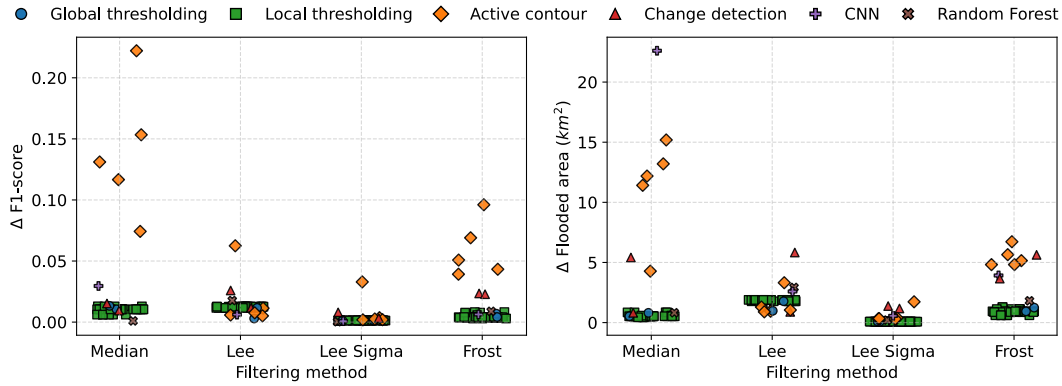


Figure 11: Range of variation (between the minimum and maximum values) of F1-score and flooded area due to filtering methods hyperparameters for the 48 flood mapping configurations for 2 February 2021 acquisition. The points are jittered horizontally for visualization purposes.

4.3.3 Impact of hyperparameters

The role of hyperparameters used in flood mapping methods was evaluated, excluding the supervised classification methods (CNN and Random Forest), for which no hyperparameters were defined in this study. Similarly to the preprocessing analysis, we fixed the preprocessing configuration to isolate and evaluate the impact of flood mapping hyperparameters.

The range of variability due to the hyperparameters of flood mapping methods for all 26 speckle filtering configurations (3 configurations for Median and Lee filters, nine configurations for Lee Sigma and Frost Filters, one configuration for SAR2SAR and no preprocessing) is evaluated. The results for the acquisition on 2 February 2021 are presented in Fig. 12. In this study, the active contour method showed the highest sensitivity to flood mapping hyperparameter

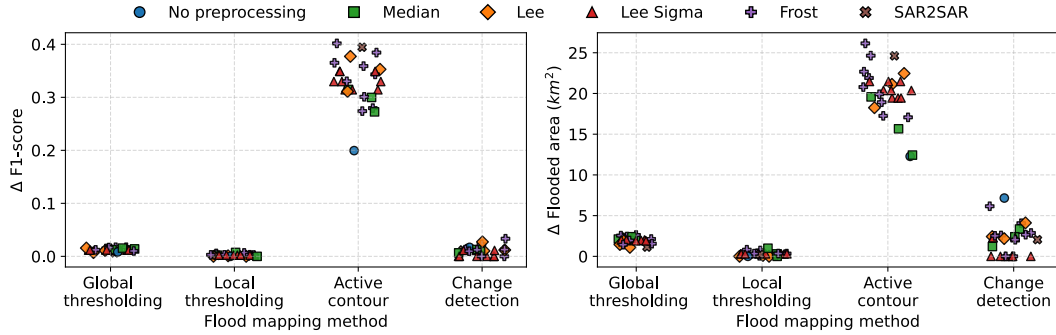


Figure 12: Range of variation (between the minimum and maximum values) of F1-score and flooded area due to flood mapping methods hyperparameters for the 26 speckle filtering configurations for 2 February 2021 acquisition. The points are jittered horizontally for visualization purposes.

variations, with F1-score differences reaching up to 0.4 and differences in flooded areas ranging between 15 and 25 km² for all speckle filtering configurations. In contrast, change detection and local thresholding methods demonstrated low sensitivity, with F1-score differences of less than 0.03 and flooded area variations limited to a few square kilometers. For active contour models, most of the variability can be attributed to the flood mapping hyperparameters. While thresholding methods showed limited variability, both active contour and change detection methods displayed greater sensitivity depending on the fixed speckle filtering configuration. For instance, variations in F1-score or flooded area due to hyperparameters for change detection were smaller for Lee Sigma configurations compared to Median filtering configurations.

4.3.4 Impact of morphological operations

Figure 13 shows the improvement or degradation in the F1-score for global thresholding, local thresholding, and change detection methods for one of the acquisitions (2 February 2021) when using morphological operations. The color intensity represents the average gain (or loss) relative to the configurations without morphological post-processing. The active contour and supervised classification approaches are not reported here, as morphological operations had negligible or no effect on the generated flood maps due to the minimal presence of holes or small unconnected flood elements in these flood maps.

We observed that the F1-score increased for the three methods, especially for large hole filling (50 to 100 pixels) and small patch removal (10 to 50 pixels). This is mainly due to eliminating isolated false positives and filling small gaps, improving the spatial coherence of flood maps. However, with the increasing size of patch removal, the performance decreases, suggesting that the small-scale features are well-captured and should not be removed. While the role of morphological operations was evaluated, the relative improvements in the F1-score metric are relatively low. While the morphological operations improved the statistics for the F1-score, the physical representation of the flood may not be improved. Indeed, they can remove small-scale features or introduce incorrect flood pixels in non-flooded regions. Therefore, careful tuning and verification of experts for morphological operations is necessary to avoid non physical observations.

Finally, Fig. 14 compares the flood maps with the highest F1-scores obtained for each method, including those enhanced with morphological operations for 2 February 2021. The CNN-based approach achieved the best overall performance, with an F1-score of 0.769, followed by the Change Detection and Random Forest methods. The flood map produced by the CNN is the most continuous but tends to overestimate the flooded area. The active contour model also exhibits a continuous flood extent, but it significantly underestimates inundation in both the upstream and downstream regions. The other methods result in less smoothed flood extent, capturing more localized details and irregularities. While some of these details may correspond to actual observations, their fragmented nature results in poorer alignment with the continuous structure of the simulated flood map, while having pattern that match the simulated flood map closer than for CNN.

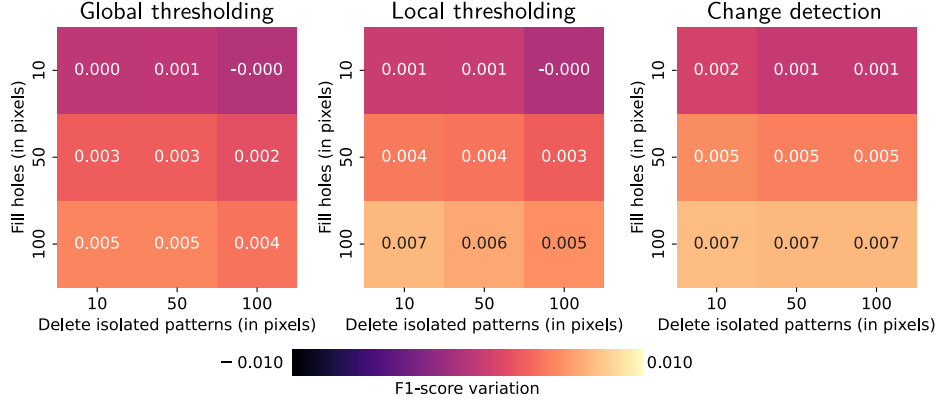


Figure 13: Heatmaps representing the average gain (or loss) in F1-score due to morphological operations relative to configurations without morphological post-processing.

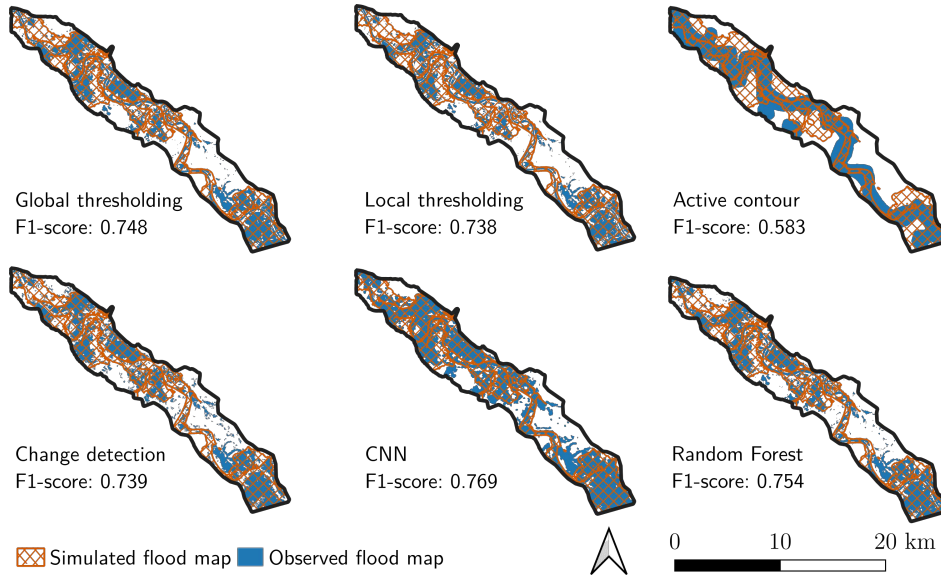


Figure 14: Comparison of the simulated flood maps (orange) with the generated flood maps (blue) for 2 February 2021.

5 Water depth estimation

The flood maps extracted from the preprocessed SAR images can be used with Digital Elevation Models (DEM) including the bathymetry of the river channel, to construct water depth fields. In the literature, the approach of Hostache et al. (2009) for water depth estimation has proved effective but requires prior expertise in the flow motions and extensive fieldwork to constrain the method. In this study, we preferred to compare automatic water depth estimation methods on the whole flood map or specific cross sections. Methods such as the Flood Water Depth Estimation Tool (Fw-DET) (Cohen et al., 2018) or the FLEXTH methodology (Betterle and Salamon, 2024) are available to derive water depth fields across the entire domain based on a flood map and DEM. Other methods rely on cross-section analysis by considering that the free surface is flat on a cross-section, thus retrieving the water depth field by knowing the locations of the edges of the flood and topography. We generated water depth fields by applying these methods across a range of hyperparameters, as summarized in Table 1. The water depth estimation workflow follows a structured sequence illustrated in Fig. 15:

1. **Input:** Flood maps derived from SAR imagery and a Digital Elevation Model (DEM), all projected on the same spatial grid. The DEM is a combination of topographic data at 1 m spatial resolution with vertical accuracy between 0.2 and 0.5 m, and the bathymetry of the channel (based on 70 cross-sections measurements).
2. **Methods:** Fw-DET, FLEXTH, or cross-section analysis. Each method is evaluated for different hyperparameters.
3. **Water depth estimation:** Each method determines the water surface elevation and subtracts the underlying terrain elevation from the DEM to generate a water depth field.

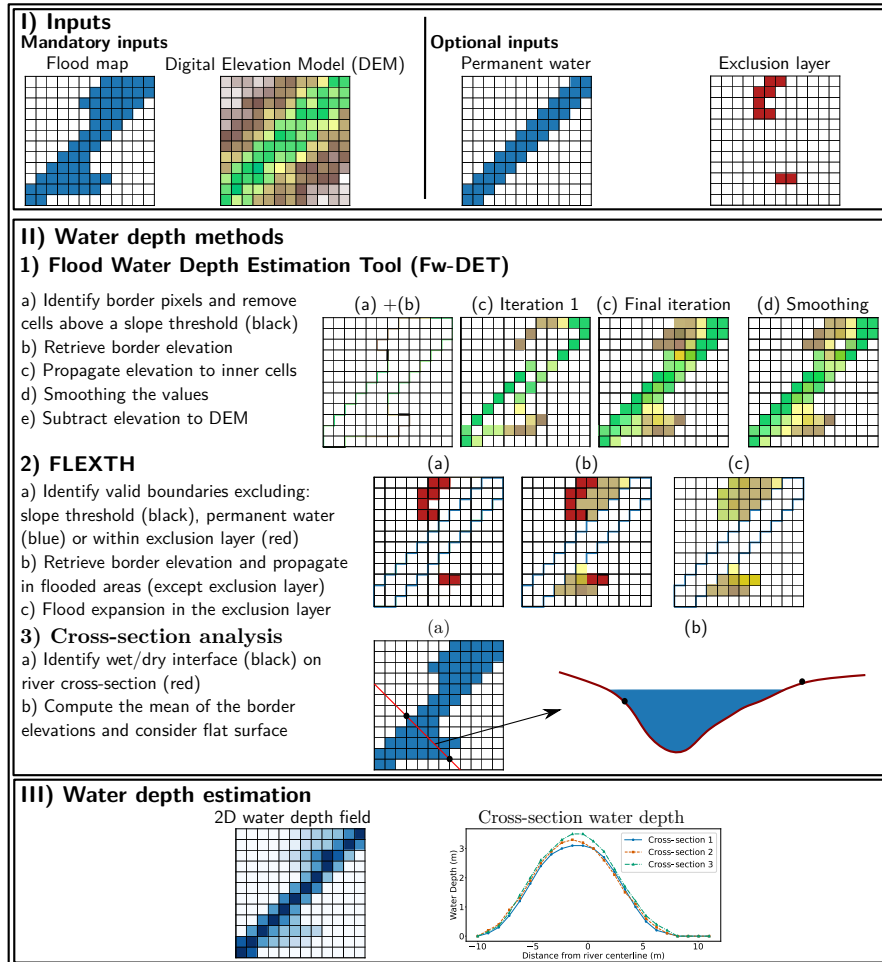


Figure 15: Schematic overview of the water depth estimation methodology.

5.1 Methods

5.1.1 Fw-DET method

The Flood Water Depth Estimation Tool (Fw-DET) (Cohen et al., 2018, 2019) quantifies water depth continuously in the domain using a flood extent polygon and a Digital Elevation Model. A schematic representation of Fw-DET principle is presented in panel II)1) of Fig. 15. Flood boundaries are derived from the flood map and converted into a line layer, while steep slope cells are filtered out using a slope threshold. This line layer is rasterized to align with the DEM grid. Elevation values are extracted for these boundary cells, and each cell within the flood extent polygon is assigned the elevation of the nearest boundary cell under the assumption of a flat water surface. Water depth is computed by

subtracting the DEM elevation from the water surface elevation. To mitigate artifacts caused by mismatches between the DEM and the flood extent, a smoothing procedure is applied multiple times using a 3×3 window.

This study considered the number of smoothing iterations and the slope threshold as hyperparameters. The latest version of the code developed by Cohen et al. (2019) is available at <https://github.com/csdms-contrib/fwdet>. For the present study, the code was adapted to operate without the QGIS interface, using standalone Python scripts to enable batch processing.

5.1.2 FLEXTH method

The FLEXTH method (Betterle and Salamon, 2024) presents an approach similar to the Fw-DET methodology but introduces improvements to mitigate unrealistic water depth estimates (Cohen et al., 2019). Figure 15, panel II(2), depicts the FLEXTH methodology schematically. The methodology was extended to account for additional inputs, such as exclusion or permanent / seasonal water body masks. The method expands the flooded area into adjacent no-data regions and estimates the water depths based on the DEM (Betterle and Salamon, 2024). As in Fw-DET, pixels along the flood boundaries with a slope exceeding a user-defined threshold are excluded from the boundary. Water depths within the flooded area are then estimated using a weighted average of the boundary cell elevations based on the closest cells up to a specified maximum number of neighbors.

The slope threshold and the maximum number of neighbors used in the computation were considered hyperparameters. In Betterle and Salamon (2024), they highlighted that these two parameters are the most influential in the FLEXTH method. The FLEXTH method, initially developed by Betterle and Salamon (2024), is available as a Python code at <https://code.europa.eu/floods/floods-river/flexth> and was adapted for the needs of our study.

5.2 Cross-section approach

The two previous methods estimated the water depths continuously across the domain. An alternative approach is to perform cross-section analysis along the river channel by extracting the surface elevation at the dry/flood interface along predefined cross-sections. The water surface is assumed to be flat for each cross-section, and its elevation is computed as the mean elevation of the identified boundary cells. However, this method is sensitive to errors caused by over-detection of the wet/dry interface along the cross-sections. To address this issue, an alternative strategy involves considering only the left and right banks of the flood extent (Schumann et al., 2007), excluding other points within the flooded area. This strategy is displayed schematically in Section II(3) of Fig. 15. In this study, we tested this second approach. The cross-section approach did not use any hyperparameters in this study.

5.3 Results

The methods for estimating water depth fields were applied to the flood maps generated in Section 4. For each satellite acquisition, more than 10,000 flood maps were generated. In our implementation, the processing time to generate the water depth field for each flood map was approximately 15 seconds for Fw-DET, 30 seconds to 1 minute for FLEXTH, and a few seconds for the cross-section approach. All computations were conducted on an Intel® Core™ i7-11850H CPU @ 2.50GHz. Thus, generating water depth fields for all hyperparameter combinations (nine configurations of the Fw-DET method, nine configurations of the FLEXTH method, and one for the cross-section approach) would result in an impractically large number of outputs and computation time. Since the generation of water depth fields depends on the input flood contour, our focus is on analyzing the variations of the water depth fields due to variations in the flood contour (due to speckle filtering, flood mapping methods, and hyperparameters used previously). We selected 10 flood maps per flood mapping method to capture a representative range of potential flood map inputs. These were sampled uniformly across the full span of their F1-scores, from the lowest to the highest score, to ensure coverage of both high- and low-quality segmentation results. The Fw-DET, FLEXTH, and cross-section methods were applied for these flood maps. For the cross-section approach, the results were analyzed for two user-defined cross-sections shown in Fig. 16. The estimated water depths were compared with the hydrodynamic simulations and watermarks presented in Section 2.1.2. The watermarks were used only for the 16 December 2019 and 3 February 2021 acquisitions (near the flood's peak). For the comparison against simulations, the Root Mean Square Error (RMSE) was computed, comparing pixel-to-pixel values. The rasterized results were projected at the measurements' locations for the watermarks. Here, contrary to the flood mapping evaluation, all the domain is taken into account including the minor bed and the exclusion zones since the three water depth estimation methods aim to estimate the water depth even in exclusion zones with specific behaviors (e.g., propagation of the water in the exclusion zone with the FLEXTH method, or with the cross-section method which considers a flat water surface between the two edges).

5.3.1 Water depth estimation methods evaluation

Figure 17 shows the RMSE of the estimated water depths compared to hydrodynamic simulations (see Fig. 17a) and watermarks (see Fig. 17) for the FwDET and FLEXTH methods. Each box plot indicates the variability explained by the input flood map or water depth estimation hyperparameters.

FLEXTH and Fw-DET demonstrate comparable performance in terms of median RMSE. Both methods exhibit a median RMSE ranging from 0.8 to 1.9 m, depending on the satellite image acquisition. The variability of the RMSE is also comparable for both algorithms, with slightly more outliers for the FLEXTH method, which may be caused by the used hyperparameter values. These conclusions are similar either with the simulated water depth fields or watermarks as validation dataset. For the watermarks, the variability is more important. The observed variability and elevated RMSE values can be attributed to the flood map selection from previous processing steps, including high and low F1-score maps relative to the reference. This variability and the influence of water depth estimation method hyperparameters are further analyzed in Section 5.3.2 for the Fw-DET and FLEXTH methods.

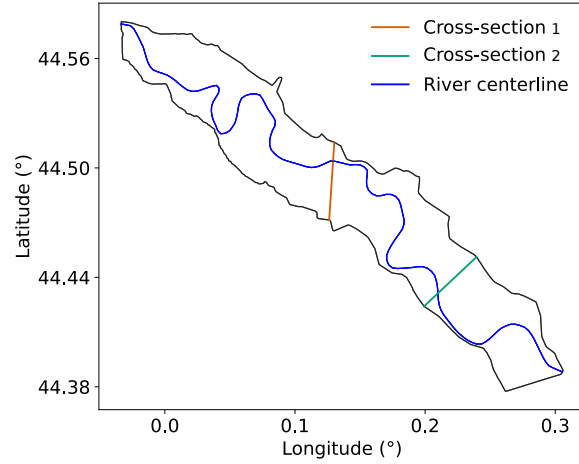


Figure 16: Visualization of the two cross-sections (cross-section 1 in orange and cross-section 2 in green) used in the study.

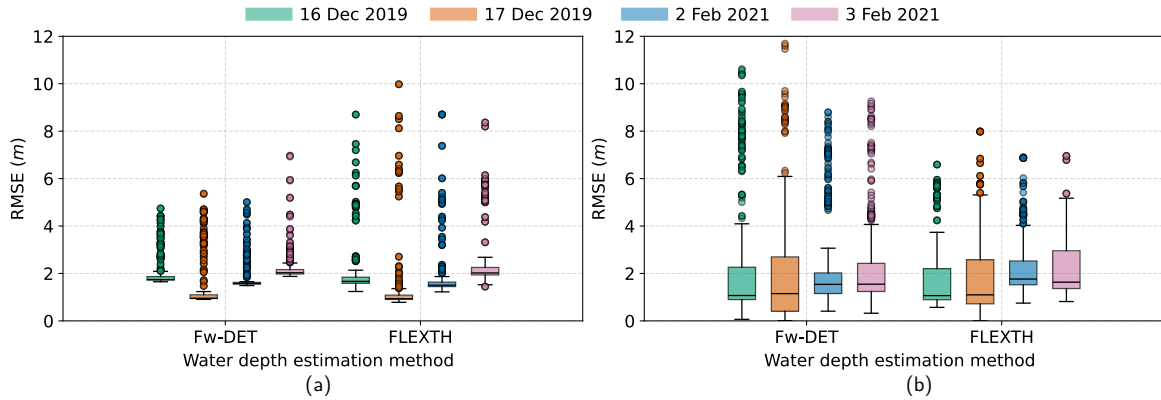


Figure 17: Root Mean Square Error (RMSE) between the estimated water depths and simulated (a) or watermarks (b) for Fw-DET and FLEXTH methods for four satellite images.

Next, Fig. 18 presents the RMSE of water depth estimates on the two selected cross-sections (see Fig. 16) for the cross-section approach, along with Fw-DET and FLEXTH methods projected onto these cross-sections. The Fw-DET and FLEXTH methods exhibited similar RMSE performance for both cross-sections, though FLEXTH showed higher variability. The cross-section approach, while showing similar or slightly higher median RMSE values compared to FLEXTH and Fw-DET on the two cross-sections, exhibited more variability, particularly for the 2 February 2021 image,

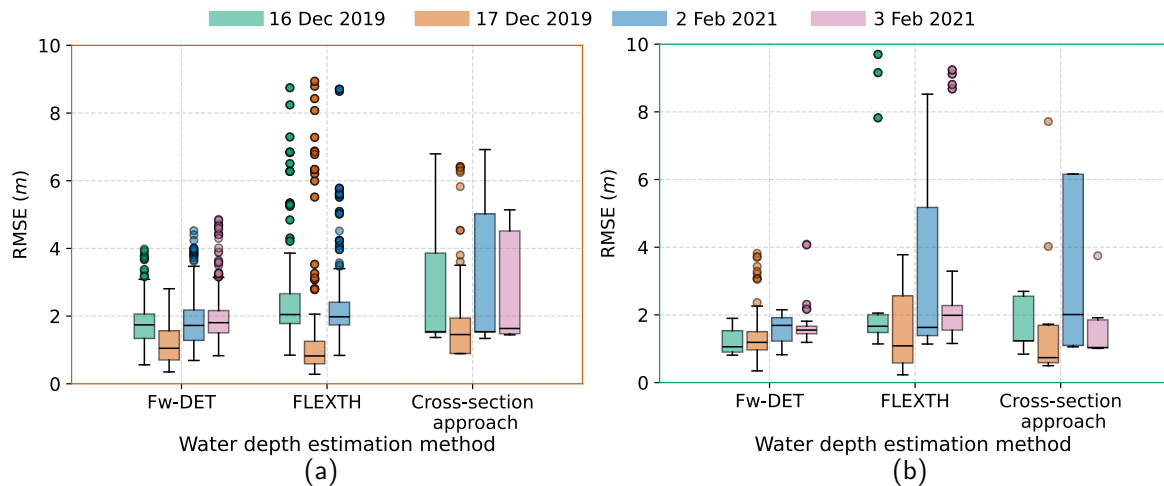


Figure 18: Comparison of Fw-DET, FLEXTH, and cross-section approaches on cross-section 1 (a) and cross-section 2 (b) for four satellite acquisitions against hydrodynamic simulations.

which was likely related to challenging flood contour detection conditions for that image. The Fw-DET and FLEXTH algorithms propagate the surface elevation from the edges of the flood to the inner flood. The cross-section approach, on the other hand, relies only on the identification of the right and left edges of the flood. Then, the cross-section approach is highly sensitive to identifying the border of the flood on the cross-section, while it is less the case for Fw-DET and FLEXTH, which rely on the whole flood extent borders. The variability of the RMSE for the cross-section approach was entirely due to the variability of the input flood maps, since the cross-section approach was considered without hyperparameters.

5.3.2 Impact of hyperparameters and flood map inputs

The influence of flood map inputs and method hyperparameters on water depth estimation performance was evaluated for the Fw-DET and FLEXTH methods. Figure 19 presents the distribution of RMSE values, computed against hydrodynamic simulations across four satellite acquisitions, grouped by flood mapping method used as input to the water depth estimation methodology.

The results indicated that the flood map used as input for the water depth estimation process significantly influenced the estimation accuracy. Flood maps derived from global and local thresholding yielded lower RMSE and variability. In contrast, flood maps produced with CNN and Active contour models lead to greater variability and higher RMSE in water depth estimates. However, the median RMSE values for CNN and active contour flood mapping methods are close to the other algorithms, suggesting that they can achieve comparable or better results with water depth estimation method parameter tuning. The outputs from CNN and active contour models are smoother, but their ability to precisely delineate flood boundaries is reduced due to the smoothness. This can result in significant errors in water depth estimates, particularly in areas with steep terrain. Furthermore, the large RMSE for the CNN method suggest that the extracted flooded surface is physically wrong on these cross-sections. The tendency of the CNN to overestimate the flood extent size leads to water edges in high topography gradient leading to important errors in the water depth estimation.

To study the influence of hyperparameters in the FLEXTH and Fw-DET methodologies, we proceeded similarly as in Section 4.3.3 for flood mapping methods. For each hyperparameter configuration (9 each for FLEXTH and Fw-DET) of water depth estimation methods, we computed the range between the lowest and maximum RMSE for every flood map input. Figure 20 presents a statistical analysis of the variability of this range for all the flood map input configurations for the 2 February 2021 acquisition.

While Fw-DET showed minimal variability in RMSE due to its hyperparameters alone (except for some outliers beyond the interquartile range), the FLEXTH method displays a much larger interquartile range of RMSE values. In this study, the Fw-DET method demonstrates more consistent performance. In contrast, FLEXTH, while potentially more accurate in ideal cases, was more sensitive to hyperparameter choices and could produce larger errors depending on the configuration used in each test case.

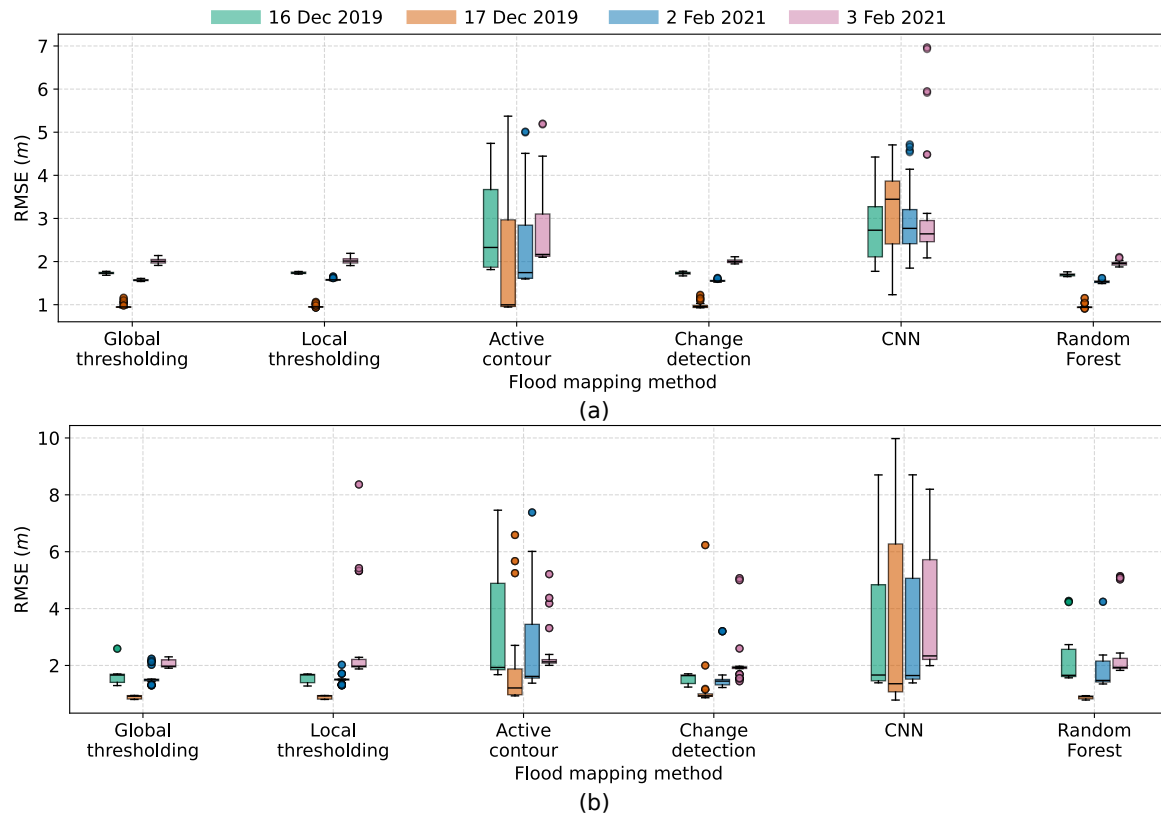


Figure 19: Impact of the flood mapping method used for water depth estimation for the Fw-DET method (a) and FLEXTH method (b) across four satellite acquisitions.

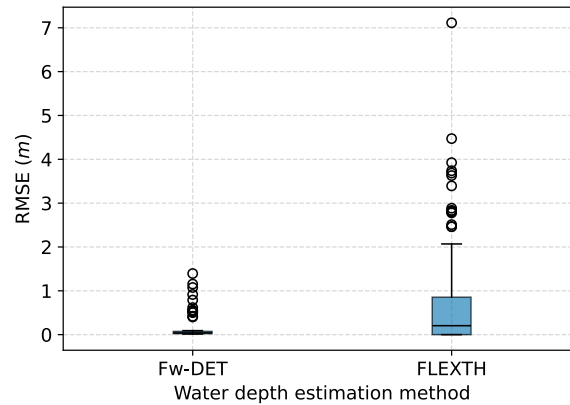


Figure 20: Box plot analysis of the variability induced only by water depth estimation hyperparameters for the 2 February 2021 acquisition.

6 Discussions

This study evaluated the influence of preprocessing, flood mapping, and water depth estimation methods using SAR imagery for hydraulic applications. Our results highlighted that each stage, from speckle filtering to method choices and hyperparameter tuning, introduced variability that can propagate through to the final outputs.

6.1 Limitations of the evaluation process

The first limitation of the study was the unavailability of ground truth data for flood maps and water depth fields in the study area. As an alternative, hydrodynamic simulations were used as a reference for evaluating performance. These simulations introduce their uncertainties due to discretization, modeled physical processes, or roughness parameterization. Then, some of the observed discrepancies between estimated and reference values may stem from errors in the hydrodynamic model rather than from the methods used to extract the information in this study. Although the availability of a reliable reference would enhance validation, the primary goal of this study was to quantify the variability introduced by preprocessing, method choices, and hyperparameter settings. However, to select the most accurate method, it is essential to use test cases where independent ground truth is available. For instance, Landuyt et al. (2018) used Copernicus Emergency Management Service flood maps extracted from SAR or optical data, and Li et al. (2018) used labeled optical images (in cloud-free conditions) to validate SAR-based flood maps. In future works, we could also consider the uncertainty of the outputs of the physics-based model in the model evaluation process, based on GLUE analysis as proposed in Huang and Merwade (2024).

In this study, hyperparameter tuning was not uniformly exhaustive across all methods. For instance, traditional flood mapping methods were evaluated across a range of hyperparameter settings, while supervised classification methods were applied using fixed configurations. This may limit their performance and underestimate their potential under optimal settings or hide their variability due to hidden hyperparameters. Although we analyzed a range of key hyperparameters, additional influential parameters may be embedded within the methods themselves. Nevertheless, we believe that the primary sources of uncertainty were adequately captured by the parameters examined in this study. Future work could expand the range of hyperparameters considered and apply uncertainty quantification techniques, such as sensitivity analysis, to quantify the influence of each parameter on the outcomes by using more exhaustive Design of Experiments. For instance, Ghosh et al. (2024) showed the impact of the choice of model architecture when using deep learning methods for flood mapping and that the importance of polarization band combinations in the performance of supervised classification. For the change detection algorithms, the only hyperparameter was the choice of thresholding method (KI or Otsu) while Tupas et al. (2023) showed that Bayes change detection could be more robust to hyperparameter changes.

The flood mapping methods applied in this methodology rely on a single polarization, although most SAR satellites, such as Sentinel-1, provide multiple polarization channels. In this study, the VV polarization was not exploited, and the flood mapping approaches could potentially benefit from the additional information it provides. However, the methods must be adapted to take into account two polarizations. Random Forest and CNN models were therefore trained using both polarizations, and the resulting models were applied to generate flood maps. Nevertheless, for the considered test case, the inclusion of an additional polarization did not lead to significant improvements in the evaluation metrics.

The intercomparison of the methods and their hyperparameterization was tested exclusively on flooded scenes. In practice, practitioners may also require methods that perform reliably under dry conditions. For the flood mapping step, supervised classification and active contour methods can be applied in both flooded and dry scenarios. In contrast, the remaining approaches rely on the automatic detection of thresholds from image histograms and fail to discriminate between flooded and dry pixels when the histogram is not bimodal which is the case under dry conditions. Consequently, supervised classification methods or alternative implementations of change detection approaches (Tupas et al., 2023) may be preferred, as the change detection method used in this work depends on the automatic identification of a threshold in bimodal histograms, a condition that is generally only met during flood events.

6.2 Hydraulic analysis and implications for flood monitoring

The results of this study highlight that the quality of flood maps and water depths estimates is highly sensitive to preprocessing choices, method selection, and hyperparameter tuning. If the uncertainty in the generated output is high, the derived products may propagate errors into the hydrodynamic models, leading to inaccurate parameter estimation. For instance, Fig. 9 highlights that most configurations led to an underestimation of the flooded area compared to simulations. In this work, a fixed parameterization was used, but in a calibration context, if the variability due to flood map processing is more important than the changes in simulated flood maps due to the roughness parameter, the calibration could be complicated. Ideally, the variability in the observations should be smaller than the variability in the simulations.

To analyze the variability of the simulated flood maps due to the model parameterization, we evaluated the hydrodynamic model for the 2021 flood event using 2,000 parameter samples under three different Strickler roughness configurations described in Table 2.

Figure 21 compares the 95% variability range in simulated flooded area for each DoE with the flooded areas obtained using various flood mapping configurations for the February 2, 2021, SAR acquisition. The variability in the simulated

Table 2: Three tested Design of Experiments (DoE) for Strickler parameterization for numerical model evaluation. DoE 1 is the baseline with fixed channel values and narrow ranges in floodplains, DoE 2 accounts for wider ranges in floodplains, and DoE 3 also considers a varying Strickler across the three channel subdomains.

Land use	Strickler ($m^{1/3} \cdot s^{-1}$): DoE 1	Strickler ($m^{1/3} \cdot s^{-1}$): DoE 2	Strickler ($m^{1/3} \cdot s^{-1}$): DoE 3
Main channel	Fixed	Fixed	$\mathcal{U}(30, 50)$
Waterbodies	$\mathcal{U}(32.5, 37.5)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$
Fields and meadows without crops	$\mathcal{U}(17.5, 22.5)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$
Cultivated fields with low vegetation	$\mathcal{U}(15, 20)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$
Cultivated fields with high vegetation	$\mathcal{U}(10, 15)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$
Shrublands and undergrowth areas	$\mathcal{U}(8, 12)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$
Areas of low urbanization	$\mathcal{U}(8, 10)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$
Highly urbanized areas	$\mathcal{U}(5, 8)$	$\mathcal{U}(5, 50)$	$\mathcal{U}(5, 50)$

flooded area is limited for DoE 1 and was lower than the variability across SAR-derived flood mapping methods. This complicates calibration because uncertainty from SAR processing dominates over that from model parameters. For the second DoE, the larger variation in floodplain roughness resulted in a wider range of flooded areas (approximately 5 km²), which is comparable to the variability observed for each SAR method, except for the active contour method. For DoE 3, there is a large spread in simulated flood extents, which is similar to the spread between methods (for instance, between the change detection median and the CNN median). For this study, only the CNN flood mapping method intersected with the simulation for DoE 1 and the median for the thresholding, change detection and Random Forest methods are close to the lower range of DoE 3. While this comparison on the flooded area can help to calibrate the model and identify subdomains where the Strickler should be adjusted, the uncertainty of the extracted flood maps should be taken into account during a calibration process. For instance, we mentioned that the CNN tend to overestimate the flooded surface even if it seems closer to the simulations. Thus it is necessary to have a relevant evaluation methodology to compare the simulations and observations given their uncertainties as discussed in Travert et al. (2025) and Landwehr et al. (2024). Using data assimilation, both of the uncertainty of the generated flood maps and the uncertain numerical model (due to the uncertain Strickler parameters) can be taken into account to optimize the parameter values. Furthermore, when calibrating flood numerical models, given the uncertainty of generated flood maps, the modeler should be cautious and consider an ensemble of flood maps or use the most unfavorable test case to estimate conservative Strickler values and avoid non-conservative simulations.

Water depth estimation methods produced water depth fields with RMSE errors of around 2 m (for the median). These errors are too significant to be ignored in the calibration of this study area. Then, the calibration should be spatially constrained to areas where water depth estimates are more reliable, such as zones with low topographic gradient.

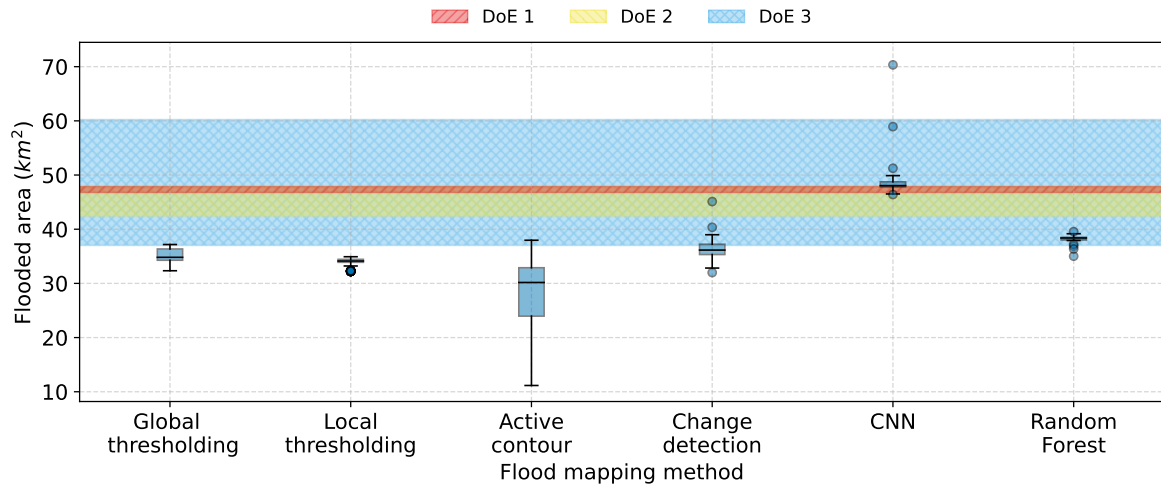


Figure 21: Comparison of flooded area variability from hydraulic simulations using three Design of Experiment (DoE 1–3) against SAR-based flood maps. The colored surface for the three DoE indicates the 95% confidence interval of simulated flooded areas.

7 Conclusions

By systematically evaluating the impact of preprocessing, flood mapping, and water depth estimation methods and their hyperparameters, this study aimed to inform best practices for generating flood maps and water depth fields from SAR imagery. It contributes to improving the reliability of SAR-based hydraulic applications by underscoring the importance of assessing the influence of speckle filtering and hyperparameter choices. Rather than relying on a single fixed configuration, the findings highlight the need for a comprehensive evaluation of processing options. These options can significantly impact the extracted flooded areas or water depths, ultimately influencing the calibration and validation of hydraulic models. The choice of speckle filtering method was shown to significantly influence the statistics of SAR images, with deep learning methods such as the SAR2SAR approach outperforming the traditional approaches. For flood mapping, supervised classification methods demonstrated the highest median accuracy, F1-score and closest flooded area, but overestimating the flood extent size and leading to important errors in the water depth estimation step. Unsupervised methods such as thresholding and change detection offered reliable results with minimal tuning, making them suitable for rapid deployment. We highlighted the role of the speckle filter method choice and its hyperparameter on the flooded area estimates, with variations in flooded area estimates. The hyperparameters of the flood mapping methods were also influential on the generated flood maps, especially for active contour models. Three methods for estimating water depth fields were compared, showing that the accuracy of water depth estimation is strongly influenced by both the quality of the input flood maps and the parameterization of the methods. In this study, and for the tested hyperparameters, FLEXTM and Fw-DET exhibited similar performance. For studying the water depth on cross-sections, the cross-section analysis results in similar estimates but with a high variability depending on the cross-sections or input flood maps.

In future flood studies using SAR imagery for hydraulic applications, the role of preprocessing, method choices, and hyperparameter tuning should be evaluated to account for the errors due to the processing pipeline. We studied only SAR images, but similar methodologies to evaluate the impact of method choices and hyperparameters could be applied to optical satellite images.

Code and data availability

The openTELEMAC software is freely available on a Gitlab server with a track of all the developments and a fixed branch for the v8p5 version used in the present study (https://gitlab.pam-retd.fr/otm/telemac-mascaret/-/tree/v8p5r0?ref_type=tags). The result files of the simulations and their projections on raster grids are provided for easier reproducibility. The code and all the data used in this study are available online at <https://github.com/jtravert/sar-flood-evaluation-framework> with some instructions and Jupyter notebooks to reproduce the methodology on the test case or other study areas and flood events.

Acknowledgments

This work was carried out as part of EDF R&D's MOISE-2 research project on river and coastal flood hazard assessment, whose support the authors gratefully acknowledge. This work was supported and funded by the French National Association of Research and Technology (ANRT) and EDF R&D with the Industrial Conventions for Training through REsearch (CIFRE grant agreement 2022/0972). The authors would like to gratefully acknowledge the open-source community, especially that of the openTELEMAC and of the used Python libraries (scikit-learn, SciPy, rasterio, numpy, matplotlib).

A Complements on flood mapping methods

A.1 Global thresholding

For computing the threshold distinguishing flooded from dry pixels, two methods are used:

- Otsu's method selects the threshold that maximizes the between-class variance, which separates the histogram of pixel intensities into two distinct classes. The between-class variance is defined as:

$$\sigma_B^2 = \omega_f \omega_b (\mu_f - \mu_b)^2, \quad (17)$$

where μ_f and μ_b are the mean intensities of the background and foreground classes, respectively. ω_f and ω_b are the foreground and background class fractions, respectively.

- The Kittler and Illingworth (KI) method also considers a mixture of two Gaussian distributions. The optimal threshold separates both classes by minimizing a cost function that quantifies the overlap between the

distributions and is calculated as follows:

$$J = 1 + 2 (\omega_f \log(\sigma_f) + \omega_b \log(\sigma_b)) - 2 (\omega_f \log(\omega_f) + \omega_b \log(\omega_b)), \quad (18)$$

where σ_f and σ_b are the standard deviations of the foreground and background classes, respectively.

A.2 Local thresholding

The three conditions to keep a tile in the local thresholding approach are evaluated with three quantitative coefficients:

- Ashman's D coefficient (Ashman et al., 1994), which quantifies the separation between two Gaussian distributions, and is defined as:

$$AD = \frac{|\mu_f - \mu_b|}{\sqrt{0.5(\sigma_f^2 + \sigma_b^2)}}. \quad (19)$$

- Bhattacharyya Coefficient (Bhattacharyya, 1943), which quantifies the amount of overlap between two distributions, and is computed as:

$$BC = \sum_k \sqrt{hist(y_k)} \sqrt{hist_f(y_k)}, \quad (20)$$

where y is the pixel values in the tiles, $hist$ is the histogram of the distribution, and $hist_f$ is the fitted histogram with two Gaussian curves, and k is the bins of the two discrete histograms.

- The surface ratio is defined as the ratio of the area (measured in number of pixels) covered by the smaller population (e.g., dry or flooded pixels) to that of the larger population.

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631.
- Ashman, K. M., Bird, C. M., and Zepf, S. E. (1994). Detecting bimodality in astronomical datasets. *arXiv preprint astro-ph/9408030*.
- Bates, P. D. (2012). Integrating remote sensing data with flood inundation models: how far have we got? *Hydrological processes*, 26(16):2515–2521.
- Bazi, Y., Bruzzone, L., and Melgani, F. (2005). An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):874–887.
- Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16):4345–4378.
- Besnard, A. and Goutal, N. (2011). Comparaison de modèles 1d à casiers et 2d pour la modélisation hydraulique d'une plaine d'inondation—cas de la garonne entre tonneins et la réole. *La Houille Blanche*, (3):42–47.
- Betterle, A. and Salamon, P. (2024). Water depth estimate and flood extent enhancement for satellite-based inundation maps. *Natural Hazards and Earth System Sciences*, 24(8):2817–2836.
- Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, 35:99–110.
- Bonafilia, D., Tellman, B., Anderson, T., and Issenberg, E. (2020). Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 210–211.
- Bovolo, F. and Bruzzone, L. (2007). A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1658–1670.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Brown, K. M., Hambidge, C. H., and Brownett, J. M. (2016). Progress in operational flood mapping using satellite synthetic aperture radar (sar) and airborne light detection and ranging (lidar) data. *Progress in Physical Geography*, 40(2):196–214.

- Bruniquel, J. and Lopes, A. (1997). Multi-variate optimal speckle reduction in sar imagery. *International journal of remote sensing*, 18(3):603–627.
- Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277.
- Chini, M., Hostache, R., Giustarini, L., and Matgen, P. (2017). A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):6975–6988.
- Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*. McGraw-Hill Book Company.
- Clement, M. A., Kilsby, C., and Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, 11(2):152–168.
- Cohen, S., Brakenridge, G. R., Kettner, A., Bates, B., Nelson, J., McDonald, R., Huang, Y.-F., Munasinghe, D., and Zhang, J. (2018). Estimating floodwater depths from flood inundation maps and topography. *JAWRA Journal of the American Water Resources Association*, 54(4):847–858.
- Cohen, S., Raney, A., Munasinghe, D., Loftis, J. D., Molthan, A., Bell, J., Rogers, L., Galantowicz, J., Brakenridge, G. R., Kettner, A. J., et al. (2019). The floodwater depth estimation tool (fwdet v2. 0) for improved remote sensing analysis of coastal flooding. *Natural Hazards and Earth System Sciences*, 19(9):2053–2065.
- Dalsasso, E., Denis, L., and Tupin, F. (2021). Sar2sar: A semi-supervised despeckling algorithm for sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4321–4329.
- Dasgupta, A., Hostache, R., Ramsankaran, R., Schumann, G. J.-P., Grimaldi, S., Pauwels, V. R., and Walker, J. P. (2021). A mutual information-based likelihood function for particle filter flood extent assimilation. *Water Resources Research*, 57(2):e2020WR027859.
- Deledalle, C.-A., Denis, L., Tupin, F., Reigber, A., and Jäger, M. (2014). Nl-sar: A unified nonlocal framework for resolution-preserving (pol)(in) sar denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2021–2038.
- Di Baldassarre, G., Schumann, G., and Bates, P. D. (2009). A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. *Journal of Hydrology*, 367(3-4):276–282.
- El Garroussi, S., de Lozzo, M., Ricci, S., Lucor, D., Goutal, N., Goeury, C., and Boyaval, S. (2019). Uncertainty quantification in a two-dimensional river hydraulic model. In *International Conference on Uncertainty Quantification in Computational Sciences and Engineering*.
- Frost, V. S., Stiles, J. A., Shanmugan, K. S., and Holtzman, J. C. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on pattern analysis and machine intelligence*, (2):157–166.
- Ghosh, B., Garg, S., Motagh, M., and Martinis, S. (2024). Automatic flood detection from sentinel-1 data using a nested unet model and a nasa benchmark dataset. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 92(1):1–18.
- Giustarini, L., Hostache, R., Matgen, P., Schumann, G. J.-P., Bates, P. D., and Mason, D. C. (2012). A change detection approach to flood mapping in urban areas using terrasar-x. *IEEE transactions on Geoscience and Remote Sensing*, 51(4):2417–2430.
- Giustarini, L., Matgen, P., Hostache, R., Montanari, M., Plaza, D., Pauwels, V., De Lannoy, G., De Keyser, R., Pfister, L., Hoffmann, L., et al. (2011). Assimilating sar-derived water level data into a hydraulic model: a case study. *Hydrology and Earth System Sciences*, 15(7):2349–2365.
- Goodman, J. W. (1976). Some fundamental properties of speckle. *JOSA*, 66(11):1145–1150.
- Grimaldi, S., Li, Y., Pauwels, V., and Walker, J. P. (2016). Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges. *Surveys in Geophysics*, 37(5):977–1034.
- Henry, J.-B., Chastanet, P., Fellah, K., and Desnos, Y.-L. (2006). Envisat multi-polarized asar data for flood mapping. *International Journal of Remote Sensing*, 27(10):1921–1929.
- Hervouet, J.-M. (2007). *Hydrodynamics of free surface flows: modelling with the finite element method*. John Wiley & Sons.
- Horritt, M. (1999). A statistical active contour model for sar image segmentation. *Image and Vision Computing*, 17(3-4):213–224.

- Hostache, R., Lai, X., Monnier, J., and Puech, C. (2010). Assimilation of spatially distributed water levels into a shallow-water flood model. part ii: Use of a remote sensing image of mosel river. *Journal of hydrology*, 390(3-4):257–268.
- Hostache, R., Matgen, P., Schumann, G., Puech, C., Hoffmann, L., and Pfister, L. (2009). Water level estimation and reduction of hydraulic model calibration uncertainties using satellite sar images of floods. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):431–441.
- Hostache, R., Matgen, P., and Wagner, W. (2012). Change detection approaches for flood extent mapping: How to select the most adequate reference image from online archives? *International journal of applied earth observation and geoinformation*, 19:205–213.
- Huang, T. and Merwade, V. (2024). Beyond a fixed number: Investigating uncertainty in popular evaluation metrics of ensemble flood modeling using bootstrapping analysis. *Journal of Flood Risk Management*, 17(2):e12982.
- Hunter, N. M. (2005). *Development and assessment of dynamic storage cell codes for flood inundation modelling*. PhD thesis, University of Bristol.
- Kittler, J. and Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition*, 19(1):41–47.
- Lai, X., Liang, Q., Yesou, H., and Daillet, S. (2014). Variational assimilation of remotely sensed flood extents using a 2-d flood model. *Hydrology and Earth System Sciences*, 18(11):4325–4339.
- Landuyt, L., Van Wesemael, A., Schumann, G. J.-P., Hostache, R., Verhoest, N. E., and Van Coillie, F. M. (2018). Flood mapping based on synthetic aperture radar: An assessment of established approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):722–739.
- Landwehr, T., Dasgupta, A., and Waske, B. (2024). Towards robust validation strategies for eo flood maps. *Remote Sensing of Environment*, 315:114439.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- Lee, J.-S. (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE transactions on pattern analysis and machine intelligence*, (2):165–168.
- Lee, J.-S. (1983). Digital image smoothing and the sigma filter. *Computer vision, graphics, and image processing*, 24(2):255–269.
- Lee, J.-S., Jurkevich, L., Dewaele, P., Wambacq, P., and Oosterlinck, A. (1994). Speckle filtering of synthetic aperture radar images: A review. *Remote sensing reviews*, 8(4):313–340.
- Lee, J.-S. and Pottier, E. (2017). *Polarimetric radar imaging: from basics to applications*. CRC press.
- Lee, J.-S., Wen, J.-H., Ainsworth, T. L., Chen, K.-S., and Chen, A. J. (2008). Improved sigma filter for speckle filtering of sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):202–213.
- Li, Y., Martinis, S., Plank, S., and Ludwig, R. (2018). An automatic change detection approach for rapid flood mapping in sentinel-1 sar data. *International journal of applied earth observation and geoinformation*, 73:123–135.
- Manning, R., Griffith, J. P., Pigot, T., and Vernon-Harcourt, L. F. (1890). *On the flow of water in open channels and pipes*.
- Martinis, S. (2010). *Automatic near real-time flood detection in high resolution X-band synthetic aperture radar satellite data using context-based classification on irregular graphs*. PhD thesis, Imu.
- Martinis, S., Groth, S., Wieland, M., Knopp, L., and Rättich, M. (2022). Towards a global seasonal and permanent reference water product from sentinel-1/2 data for improved flood mapping. *Remote Sensing of Environment*, 278:113077.
- Martinis, S., Kersten, J., and Twele, A. (2015a). A fully automated terrasar-x based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:203–212.
- Martinis, S., Kuenzer, C., Wendleder, A., Huth, J., Twele, A., Roth, A., and Dech, S. (2015b). Comparing four operational sar-based water and flood detection approaches. *International Journal of Remote Sensing*, 36(13):3519–3543.
- Martinis, S., Twele, A., and Voigt, S. (2009). Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution terrasar-x data. *Natural Hazards and Earth System Sciences*, 9(2):303–314.
- Mason, D., Schumann, G.-P., Neal, J., Garcia-Pintado, J., and Bates, P. (2012). Automatic near real-time selection of flood water levels from high resolution synthetic aperture radar images for assimilation into hydraulic models: A case study. *Remote Sensing of Environment*, 124:705–716.

- Mason, D. C., Speck, R., Devereux, B., Schumann, G. J.-P., Neal, J. C., and Bates, P. D. (2009). Flood detection in urban areas using terrasars-x. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):882–894.
- Mateo-Garcia, G., Veitch-Michaelis, J., Smith, L., Oprea, S. V., Schumann, G., Gal, Y., Baydin, A. G., and Backes, D. (2021). Towards global flood mapping onboard low cost satellites with machine learning. *Scientific reports*, 11(1):1–12.
- Montanari, M., Hostache, R., Matgen, P., Schumann, G., Pfister, L., and Hoffmann, L. (2009). Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensing-derived water stages. *Hydrology and Earth System Sciences*, 13(3):367–380.
- Morvan, H., Knight, D., Wright, N., Tang, X., and Crossley, A. (2008). The concept of roughness in fluvial hydraulics and its formulation in 1d, 2d and 3d numerical simulation models. *Journal of Hydraulic Research*, 46(2):191–208.
- Mumford, D. B. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*.
- Nguyen, T. H., Ricci, S., Piacentini, A., Fatras, C., Kettig, P., Blanchet, G., Peña Luque, S., and Baillarin, S. (2022). Dual state-parameter assimilation of sar-derived wet surface ratio for improving fluvial flood reanalysis. *Preprint*.
- Oberstadler, R., Hönsch, H., and Huth, D. (1997). Assessment of the mapping capabilities of ers-1 sar data for flood mapping: a case study in germany. *Hydrological processes*, 11(10):1415–1425.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.
- Pulvirenti, L., Marzano, F. S., Pierdicca, N., Mori, S., and Chini, M. (2013). Discrimination of water surfaces, heavy rainfall, and wet snow using cosmo-skymed observations of severe weather events. *IEEE transactions on geoscience and remote sensing*, 52(2):858–869.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18, pages 234–241. Springer.
- Schumann, G., Di Baldassarre, G., Alsdorf, D., and Bates, P. (2010). Near real-time flood wave approximation on large rivers from space: Application to the river po, italy. *Water Resources Research*, 46(5).
- Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., and Pfister, L. (2007). High-resolution 3-d flood information from radar imagery for flood hazard management. *IEEE transactions on geoscience and remote sensing*, 45(6):1715–1725.
- Schumann, G., Pappenberger, F., and Matgen, P. (2008). Estimating uncertainty associated with water stages from a single sar image. *Advances in water resources*, 31(8):1038–1047.
- Sezgin, M. and Sankur, B. I. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–168.
- Tarpanelli, A. and Benveniste, J. (2019). On the potential of altimetry and optical sensors for monitoring and forecasting river discharge and extreme flood events. In *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment*, pages 267–287. Elsevier.
- Travert, J.-P., Boyaval, S., Goeury, C., Bacchi, V., and Zaoui, F. (2025). Evaluation of performance measures for comparing flood models with satellite observations. *Water Resources Research*.
- Tupas, M. E., Roth, F., Bauer-Marschallinger, B., and Wagner, W. (2023). An intercomparison of sentinel-1 based change detection algorithms for flood mapping. *Remote Sensing*, 15(5):1200.
- Twele, A., Cao, W., Plank, S., and Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing*, 37(13):2990–3004.
- Wagner, W., Bauer-Marschallinger, B., Roth, F., Raiger-Stachl, T., Reimer, C., McCormick, N., Matgen, P., Chini, M., Li, Y., Martinis, S., et al. (2026). The fully-automatic sentinel-1 global flood monitoring service: Scientific challenges and future directions. *Remote Sensing of Environment*, 333:115108.
- Zhao, G., Bates, P., Neal, J., and Pang, B. (2020). Design flood estimation for global river networks based on machine learning models. *Hydrology and Earth System Sciences Discussions*, 2020:1–25.

4.3 Additional Case Study

The methodology presented in Section 4.2 was initially applied to a study area along the Garonne River in France. To further assess the evaluation of the processing strategy on satellite imagery, we applied the same methodology to a second study area where satellite observations and validation data are available and have already been used in previous studies [Landuyt et al., 2018]. It allows for evaluating the impact of preprocessing, method choices, and hyperparameter tuning independently of a validation dataset based on hydrodynamic simulations.

4.3.1 Methodology

The workflow for this additional case follows the same structure as described in Section 4.2 and is illustrated in Figure 4.1. The water depth estimation methods were not evaluated in this case study, as no in-situ water depth measurements were available during the flood events.

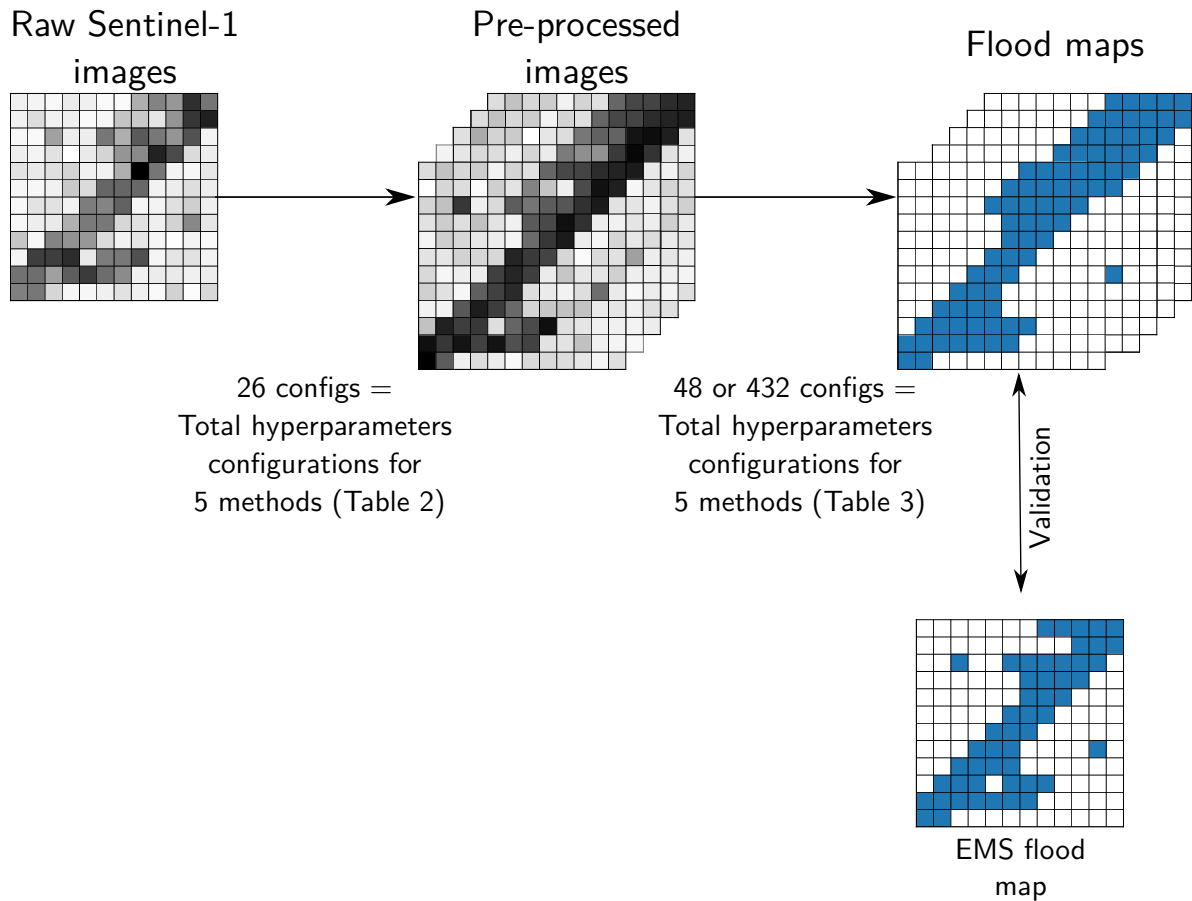


Figure 4.1: Workflow for processing Sentinel-1 (SAR) images to derive flood maps. Raw images are preprocessed using 26 hyperparameter configurations (Table 2 in Section 4.2), then each image is turned into 48 flood maps (or 432 if using morphological operations) (Table 3 in Section 4.2).

4.3.1.1 Study area

Between late 2015 and early 2016, Great Britain and Ireland experienced a series of heavy rainfall events, resulting in widespread flooding. The Shannon Basin in eastern Ireland,

located between Lough Ree and Lough Derg, was severely impacted by these floods. This study area mainly comprises peat bogs, surrounded by pastures [Landuyt et al., 2018], and is visualized in Figure 4.2.

4.3.1.2 Satellite observations

During the flood event, a Sentinel-1 image of the study area was captured on January 9, 2016, at 18:22 UTC. For the change detection method, a reference image under non-flood conditions is required. This reference was obtained from a Sentinel-1 acquisition on the same orbit on October 29, 2015, at 18:22 UTC.

4.3.1.3 Validation data

The generated flood maps were validated using reference flood maps produced by the EMS of the Copernicus programme. The EMS provides rapid mapping and monitoring services in response to natural disasters, such as floods, based on satellite and ancillary data. The flood maps were based on a COSMO-SkyMed satellite (SAR sensor) image captured in the study area on January 9, 2016, at 18:39 UTC [Copernicus Emergency Management Service, 2015]. The EMS flood map was produced using a semi-automatic approach and is shown in Figure 4.2.

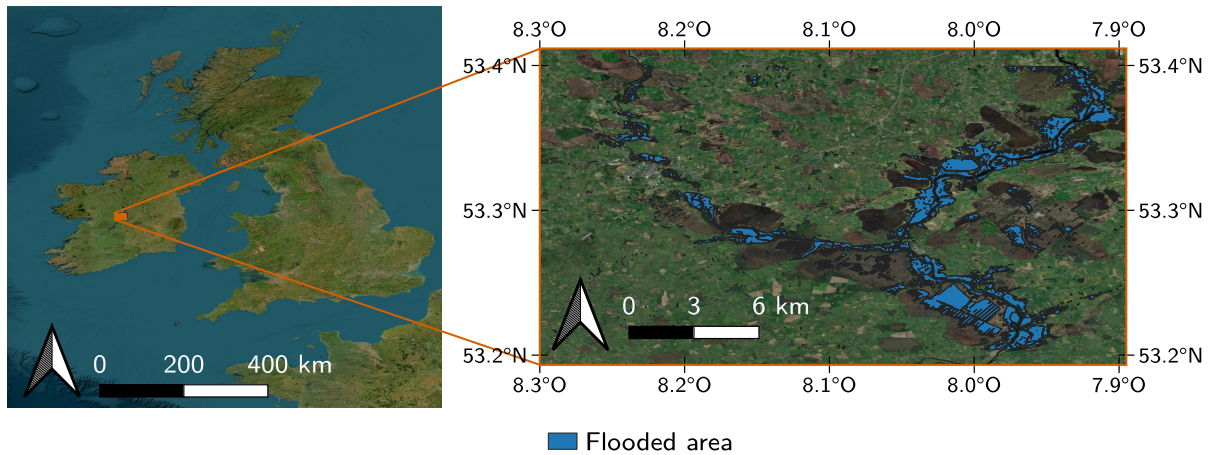


Figure 4.2: Study area and flood map produced by the Copernicus Emergency Mapping Service on the study area using a COSMO-SkyMed image acquired on January 9, 2016, at 18:39 UTC.

4.3.2 Results

Unlike the detailed analysis provided in Section 4.2, which includes a step-by-step description of the processing chain, this section focuses on summarizing the main findings from the comparison between the generated flood maps and the EMS reference map. Flood maps were generated for the 26 speckle filtering configurations and 48 flood mapping configurations, resulting in a total of 1,222 generated flood maps.

4.3.2.1 Flood mapping methods evaluation

Figure 4.3 shows the distribution of the Accuracy and F1-score values calculated between the generated flood maps and the reference satellite image described in Section 4.3.1.3.

The variability of the flooded surface for the different methods and configurations is also shown, with a dashed line representing the flooded area on the EMS flood map.

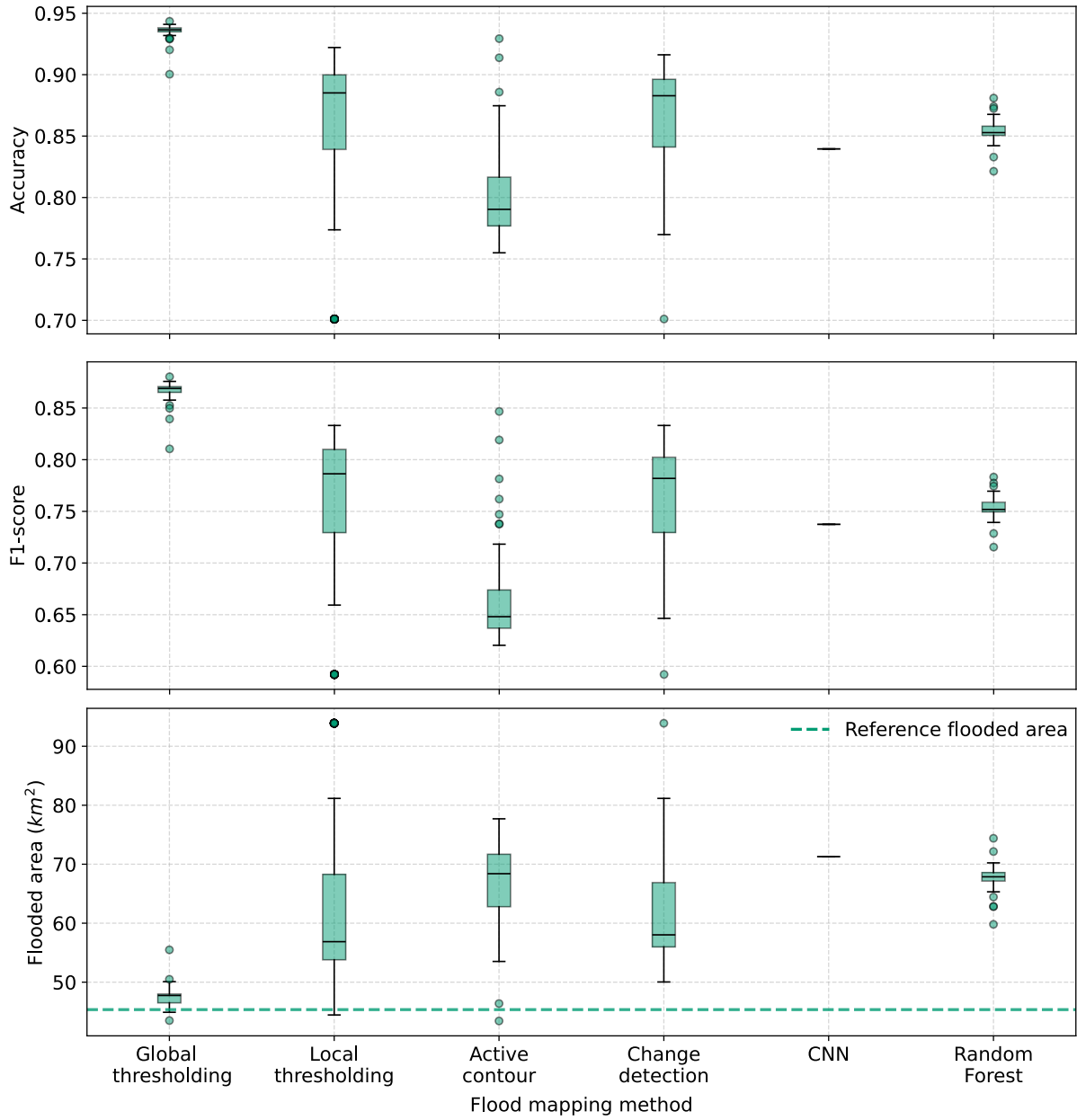


Figure 4.3: Accuracy and F1-score computed between the generated flood maps and the Emergency Mapping Service reference map grouped by flood mapping method, along with the flooded surface area for the generated flood maps.

For the Shannon test case, global thresholding shows the best performance, with median Accuracy and F1-score close to 0.94 and 0.87. It also provided a flooded surface area that aligns well with the reference. Local thresholding and change detection methods also performed well, with median Accuracy values of 0.88 and F1-scores of 0.79, but they showed higher variability, suggesting a stronger dependence on the configuration settings. Supervised classification methods (CNN and Random Forest) achieve lower median Accuracy and F1-scores compared to thresholding-based methods or change detection. These approaches overestimated the flooded area by nearly 20 km² for this test case, resulting in reduced performance despite low variability. This low variability is only due to the speckle filtering configurations, since no hyperparameters were tuned for supervised classification

methods. Active contour methods had the lowest scores among the methods, with a high variability and a large number of outliers.

Finally, the high agreement of global and local thresholding methods with the EMS reference map may be partially explained by the semi-automated nature of EMS mapping workflows, which often incorporate thresholding-like techniques. This may contribute to the strong alignment observed in both accuracy and estimated flood extent for these methods.

4.3.2.2 Impact of preprocessing

As described in Section 4.2 and illustrated in Figure 4.1, 48 flood mapping configurations were tested. For each flood mapping configuration, the different speckle filtering configurations were tested. We computed the range between the minimum and maximum values of both F1-score and estimated flooded area across these hyperparameter variations. For the 48 flood mapping configurations, Figure 4.4 shows the sensitivity of each flood mapping approach to speckle filter hyperparameters by plotting the range of variation in F1-score (left) and flooded area (right) for each configuration. Each point corresponds to one flood mapping configuration.

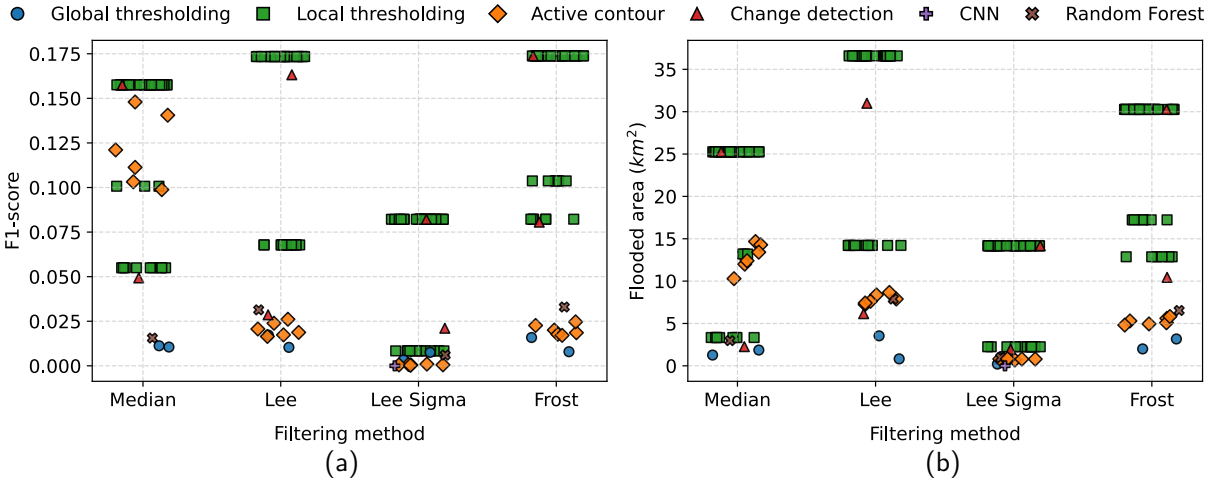


Figure 4.4: Range of variation (between the minimum and maximum values) of F1-score (a) and flooded area (b) due to filtering methods hyperparameters for the 48 flood mapping configurations. The points are jittered horizontally for visualization purposes.

The results indicate that the sensitivity to speckle filter hyperparameters varies depending on the flood mapping method and configuration. For all filtering methods, local thresholding consistently shows high sensitivity to hyperparameter settings, both in terms of F1-score and estimated flooded area. Active contour configurations are very sensitive to the window size hyperparameter of the Median filter. For the change detection configurations, one of the two configurations demonstrated a pronounced variation in performance across all filter settings, particularly with the Lee filter. In contrast, supervised classification methods (CNN and Random Forest) show limited sensitivity to filtering hyperparameters.

The Lee Sigma filter shows less spread across hyperparameter values, with a maximum variation of 4 km² for most configurations. However, variations are more important (around 15 km² for the flooded area) for one of the change detection configurations and approximately half of the local thresholding configurations.

4.3.2.3 Impact of hyperparameters

A similar analysis as in Section 4.3.2.2 is conducted to evaluate the influence of flood mapping hyperparameters on the generated flood maps. For each flood mapping method with hyperparameters (global and local thresholding, active contour, and change detection), multiple hyperparameter settings were tested as described in Section 4.2. For each of the 26 speckle filtering configurations, we computed the range between the minimum and maximum values of F1-score and flooded area resulting from the tested hyperparameter combinations. Figure 4.5 shows the variation for each filtering configuration, with each point representing the variation of a given preprocessing configuration to the flood mapping hyperparameters.

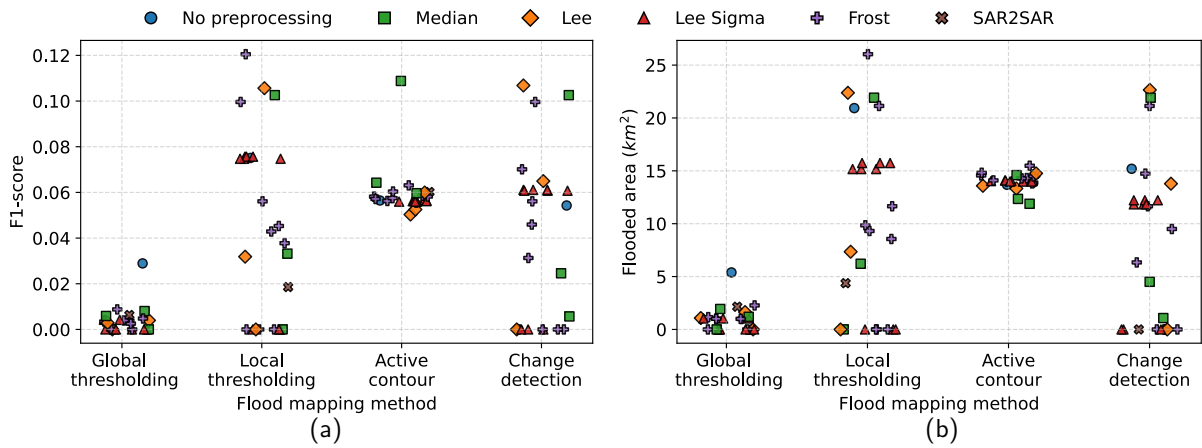


Figure 4.5: Range of variation (between the minimum and maximum values) of F1-score (a) and flooded area (b) due to flood mapping methods hyperparameters for the 26 speckle filtering configurations. The points are jittered horizontally for visualization purposes.

The results suggested that global thresholding is not very sensitive to hyperparameter variations (i.e., the method used to determine the threshold) except when no speckle filtering is applied. This suggests that global thresholding performs more consistently when preceded by speckle filtering, which stabilizes pixel intensity distributions. In contrast, the active contour method showed high sensitivity to its hyperparameters, with variations in the flooded area of around 15 km^2 and an F1-score variation of around 0.06 across most fixed speckle filtering configurations. Local thresholding and change detection methods showed variable variations across preprocessing configurations. Some speckle filtering methods result in relatively stable flood map outputs across the different hyperparameter settings, while others lead to higher variations. These findings underscore the significant influence of flood mapping hyperparameters on generated flood maps, except for the global thresholding method. They also showed that the effect of hyperparameter tuning is not uniform and depends strongly on the speckle filtering configuration. This emphasizes the need to consider both preprocessing and flood mapping method choices jointly when designing flood mapping workflows.

4.3.2.4 Comparison with the Garonne study area

The main conclusions of this additional test case are similar, indicating the high variability in flood map generation due to the choice of preprocessing, flood mapping method, and hyperparameters, with variations of 10 to 20 km^2 in the flooded area estimate depending on the chosen flood method.

However, between the two test cases, some peculiar differences and similarities are observed. For instance, in the Garonne test case, supervised methods performed better, while thresholding approaches performed better on the Shannon River. The lower performance of the supervised methods on the Shannon test case may be explained by the training data being more adapted to the flood event that occurred in the Garonne area. Furthermore, the supervised methods result in very smooth results for flood maps, while the validation flood map on the Shannon River contains holes and is fragmented (see Figure 4.2) compared to the validation flood map of the Garonne test case. Similarly to the Garonne test case, the active contour model resulted in the lowest performance and with high variability due to its hyperparameters. As for the thresholding and change detection approaches, they were robust in both test cases, and especially for the Shannon test case. This may be explained by a bias in the validation flood map, segmented using automatic approaches by the Copernicus Emergency Mapping Service.

For the effects of preprocessing, the results are consistent between the two test cases, with an important spread due to the choice of hyperparameters for the Median filter and a lower spread with variations in the hyperparameters of the Lee Sigma filter. For the Lee and Frost filters, in both studies, we highlighted the role of the hyperparameters of the filters, especially for the flood map resulting from active contour segmentation. However, on the Shannon test case, a peculiar behavior is observed for variations in speckle filter hyperparameters, with two steps and a high influence of the hyperparameter choice, whereas this was not the case in the Garonne test case. The main explanation is that to generate the flood maps in the Shannon test case, the local thresholding approach needs to consider more sub-tiles because the study area is mainly dry, with many sub-tiles not meeting the criteria defined in the method (bimodality, surface ratio).

In the Shannon test case, local thresholding and change detection methods were sensitive to variations in the hyperparameters of flood mapping methods. In the Garonne test case, the sensitivity to this hyperparameter was mainly observed for the active contour model and slightly for change detection. It shows that for each flood mapping method, the variability due to hyperparameters can vary depending on the test case.

4.4 Graphical User Interface

Hydraulic modeling requires several steps, including spatial data preparation, model construction, parameter tuning, and results visualization. For operational studies, these tasks are time-consuming and sometimes not user-friendly, particularly when using command-line scripts or a sequence of independent tools. To address this issue, many hydraulic modeling environments adopt Graphical User Interfaces to include all processing steps into a single user-friendly platform. For instance, plugins like Q4TS were designed within QGIS to embed the geospatial data preparation, pre-processing of open-TELEMAC meshes, computation launch, and visualization of results in a single environment.

In the same spirit, we developed a GUI for reproducing the workflow described in this chapter and enabling the user to have a single application, from downloading satellite data to generating final hydraulic outputs, such as flood maps or water depth fields.

The first panel of the GUI provides an interface for downloading Sentinel-1 satellite data from the Alaska Satellite Facility (ASF) APIs, and visualizing data. The user can draw a polygon or load a shapefile layer to download all available Sentinel-1 images between two dates (see Figure 4.6a). This step is not mandatory if the user has already downloaded the data. On this first panel, the user can also load any raster files and visualize them (see Figure 4.6b).

Then, three panels are available for preprocessing, flood mapping, and water depth

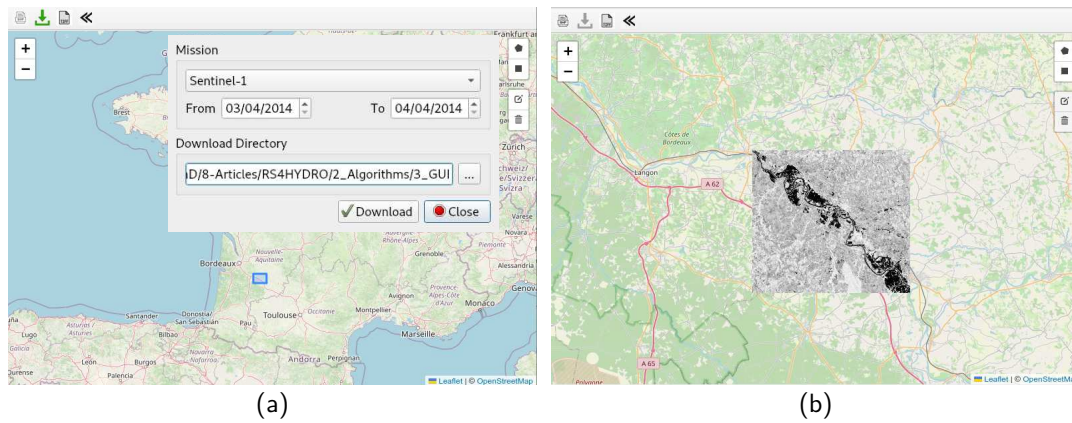


Figure 4.6: Visualization of the Graphical User Interface for Sentinel-1 data downloading (a) and satellite image visualization (b).

estimation. Figure 4.7 illustrates the GUI for preprocessing, with the configuration selected by the user on the left and a visualization of the generated products on the right. In place of the GUI, Jupyter notebooks are also available to apply the entire evaluation process.

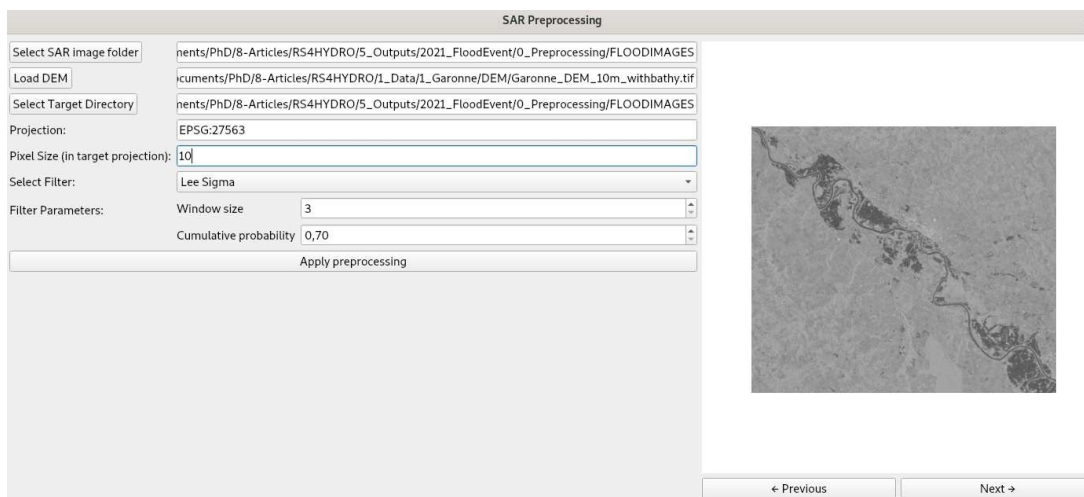


Figure 4.7: Visualization of the Graphical User Interface for the preprocessing step.

4.5 Summary

Throughout this chapter, we reviewed the influence of processing choices on the extraction of flood maps and water depth fields from SAR imagery. While most studies in the literature using satellite images for flood modeling consider unique configurations (deterministic choice of preprocessing, flood mapping, or water depth estimation methods, and hyperparameters), we demonstrated that the outputs, flood maps or water depth fields, can vary significantly depending on the processing strategy. Through two test cases (the Garonne River and the Shannon River), we showed that the best-performing method varies across regions, highlighting the need to quantify uncertainties in each study area. The results emphasize the importance of evaluating flood observation accuracy before using it for model calibration or validation. The next chapter proposes a methodology for selecting suitable performance measures to compare simulated and observed flood maps.

Chapter 5

Evaluation of Performance Measures for Comparing Flood Models with Satellite Observations

Résumé :

Ce chapitre propose une méthodologie pour guider la sélection de mesures de performance pour comparer les simulations numériques d'un événement d'inondation avec une image satellite, pour calibrer ou valider les modèles numériques. Face à un ensemble de mesures de performance, il est nécessaire d'évaluer lesquelles sont les plus appropriées pour comparer les cartes d'inondation observées et simulées. Tout d'abord, une analyse de corrélation a été réalisée pour déterminer les relations entre les différentes mesures de performance. Une série de critères a ensuite été appliquée pour évaluer la sensibilité des mesures de performance par rapport à l'étendue de l'inondation, aux caractéristiques du satellite (position et orientation) et aux erreurs de mesure ou de traitement (valeurs brutes du satellite ou extraction des cartes d'inondation). Les mesures de performance répondant au plus grand nombre de critères ont ensuite été classées en fonction de l'identifiabilité des valeurs du paramètre.

Abstract:

The present chapter proposes a methodology to guide the selection of performance measures to compare numerical simulations of a flood event with a satellite image, for purposes of model calibration or validation. Faced with a set of performance measures, practitioners need to assess which are most suitable for comparing observed and simulated flood maps. First, a correlation analysis was conducted to determine the relationships between various performance measures. A series of criteria were then applied to evaluate the sensitivity of the performance measures with respect to flood extent, satellite characteristics (position and orientation), and errors in measurement or processing (raw satellite values or extraction of flood maps). The performance measures meeting the most criteria were then ranked based on the identifiability of parameter values with regard to observation errors.

Contents

5.1	Motivations	133
5.2	Published article in Water Resources Research	133

5.1 Motivations

Satellite observations can be leveraged to extract flood maps and provide distributed data of the flood extent as discussed in Chapters 2 and 4. These observations can be used to calibrate 2D hydrodynamic models. In most automated calibration strategies such as the DA techniques presented in Chapter 3, the misfit between the numerical model and the observations is computed. In classical DA, the classical mathematical tool to compute this misfit is the Euclidean distance. However, other performance measures could be used to compute the misfit between observed and simulated flood maps which consist of dense images with geometric textures.

This chapter directly addresses the second research question introduced in Chapter 1:

Research Question 2

How can the agreement between a satellite observation and a simulation be quantified?

Section 5.2 presents the proposed methodology to guide the selection of a performance measure to compare simulated and observed flood maps. This section is written in the form of a paper which is published in Water Resources Research.

5.2 Published article in Water Resources Research

This article is available online at <https://doi.org/10.1029/2024WR038506>.

Water Resources Research®

RESEARCH ARTICLE

10.1029/2024WR038506

Key Points:

- A methodology is reported to select a performance measure; for numerical comparison between flood models and satellite observation data
- Possible performance measures are ranked according to their robustness with regard to satellite observation errors
- The methodology was tested in a real flood case study, and was effective in selecting the most relevant performance measures

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J.-P. Traver, jean-paul.travert@edf.fr

Citation:

Travert, J.-P., Boyaval, S., Goeury, C., Bacchi, V., & Zaoui, F. (2025). Evaluation of performance measures for comparing flood models with satellite observations. *Water Resources Research*, 61, e2024WR038506. <https://doi.org/10.1029/2024WR038506>

Received 26 JUL 2024
Accepted 8 MAY 2025

© 2025. The Author(s).
This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Evaluation of Performance Measures for Comparing Flood Models With Satellite Observations

J.-P. Traver^{1,2} , S. Boyaval^{2,3} , C. Goeury^{1,2} , V. Bacchi¹ , and F. Zaoui¹ 

¹EDF R&D, Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, France, ²Laboratoire d'Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France, ³Inria, Paris, France

Abstract Selecting appropriate performance measures is crucial for calibrating numerical flood models with satellite images, as differences between performance measures can significantly impact model predictions. The present study reports a step-by-step methodology to guide the selection of performance measures to compare numerical simulations of a flood event with a satellite image, for purposes of model calibration or validation. Faced with a set of performance measures, practitioners need to assess which are most suitable for comparing observed and simulated flood maps. First, a correlation analysis was conducted to determine the relationships between various performance measures. A series of criteria were then applied to evaluate the sensitivity of the performance measures with respect to flood extent, satellite characteristics (position and orientation), and errors in measurement or processing (raw satellite values or extraction of flood maps). The performance measures meeting the most criteria were then ranked based on the identifiability of parameter values with regard to observation errors. The methodology is discussed numerically in the context of a single flood event (on the Garonne River in France in February 2021), using a distribution of water depths simulated from a shallow-water model (parameterized by an uncertain friction field). The methodology was shown to be useful in the study case in identifying 4 performance measures out of 24. The proposed methodology is intended to be generic, and therefore, applicable to any flood model or flood event.

1. Introduction

To assess the state of an ongoing flood in real time or to predict likely future floods, river water motion is often simulated on numerical models. However, these models of the environment only approximate real-world physical processes. They constantly need to be corrected, by calibration to observations, to maintain forecasting capacity. Calibration is now enhanced by the development of information technology resources. Although flood models used mainly to be calibrated with pointwise gauges, it is now becoming standard practice to compare them with satellite images (Dasgupta et al., 2021; M. S. Horritt, 2000; Wood et al., 2016). The increasing availability of new types of observation, such as spatially distributed images, requires new tools for comparison between simulations and observations. In this context, which performance measures are selected for evaluating simulated versus observed flood maps significantly affects prediction (Beven, 2006).

Performance measures (also called calibration metrics, scores, verification metrics, distance, dissimilarity measures, etc.) are scalar functions that quantify differences between numerical values of a mathematical object, such as simulated and observed rasters (grids of pixels). Murphy (1993) distinguished three criteria determining simulation goodness: the match between the practitioners' judgments and the simulations, the statistical match between the simulations and observations, and the economic/environmental priorities of the decision-makers. The present paper focuses on the second issue: the quality quantification between numerical simulations and observations, based on performance measures.

The choice of an appropriate performance measure depends on the nature of the observations and/or the characteristics of the model. To calibrate the parameters of two-dimensional (2D) flood models in floodplains, it is not enough to rely solely on data from hydrometric stations along the channel. Sometimes, water mark measurements are available in floodplains (e.g., Vigicrues (<https://www.vigicrues.gouv.fr/>), a flood-monitoring service that collects water marks during floods in France) (Aronica et al., 1998; Ballesteros et al., 2011), or satellite altimetric data may be available (Domeneghetti et al., 2014). Over the last two decades, with the increasing number of remote sensing observations (aerial photographs or satellite observations), new high-resolution data have become available during floods. These data are leveraged to quantify the quality of model simulations, notably in floodplains (Di Baldassarre et al., 2009; Hall et al., 2011; M. Horritt & Bates, 2002; M. S. Horritt, 2000; Neal

et al., 2012; Wood et al., 2016). Synthetic Aperture Radar (SAR) in particular images are widely used for model quality assessment. Satellite missions such as Sentinel-1 provide free data, and SAR images enable water detection even under cloudy conditions and at night (Tarpanelli et al., 2022). From these satellite images, high-resolution flood maps can be extracted (e.g., 10 m for Sentinel-1).

In contrast to water depth measurement at single points, how to choose a performance measure for 2D flood maps is as yet unclear. For data assimilation, the simulated water depth fields are typically compared with observations in an L_2 space, considering Gaussian additive errors (Cao et al., 2019; Hostache et al., 2010). However, reconstructing water depth fields from satellite flood maps is a highly uncertain procedure (Hostache et al., 2009; Schumann et al., 2007). It would be preferable to compare flood maps directly, without introducing additional errors at the reconstruction step (Cooper et al., 2019). Reproducing identified sources of error, such as measurement error in satellite images caused by foreshortening, layover due to topographic features such as mountains, or the non-detection of flooded areas due to vegetation, is challenging (Zhao et al., 2021). Ranking performance measures according to their mathematical properties (e.g., convexity or differentiability) is not straightforward. Given the number of difficult hypotheses, we prefer to use an even more indirect method: the possibility for the practitioner to identify the optimal parameterization among a set of simulations by testing various criteria.

Weather forecasting has pioneered the comparison of spatial fields (Jolliffe & Stephenson, 2012), developing performance measures mainly for rainfall and temperature fields, but this could also be extended to hydraulics (Casati et al., 2008; Gilleland et al., 2009). To date, most flood event studies relied on comparisons between simulations and observations projected on the same grid, in a grid-to-grid (or pixel-to-pixel) approach (Aronica et al., 2002; Grimaldi et al., 2016). These performance measures provide incomplete information about simulation quality, as the comparison on a pixel-to-pixel basis does not consider location mismatch, spatial structure or field features (Casati et al., 2008; Jolliffe & Stephenson, 2012; Wilks, 2011). For example, if a simulated and observed flood have the same shape and size but are not aligned, the performance measure value may be low (Gilleland et al., 2009). Pappenberger et al. (2007) reviewed performance measures based on pixel-to-pixel comparisons. They recommended that different performance measures should be used conjointly to evaluate different flood characteristics. Stephens et al. (2014) also assessed pixel-to-pixel performance measures, and reported inconsistent performance measures behaviors depending on flood event magnitude or catchment characteristics. For validation of flood maps, Landwehr et al. (2024) proposed guidelines to take account the role of the imbalance between flooded and dry pixels in selecting a performance measure, by adapting the sampling design when evaluating the performance measure. Further evaluations of performance measures have been performed in hydrology. Cloke and Pappenberger (2008) proposed a 6-criterion approach to select new performance measures. Crochemore et al. (2015) compared expert judgment and performance measures, highlighting that they may not identify the hydrograph best matching simulations and observations in the same order.

In the hydraulics community, there are few studies exploring novel performance measures or comparing performance measures for flood maps in real flood events (Hooker et al., 2022, 2023; Landwehr et al., 2024). The present study therefore sought to define a specific methodology adapted to flood maps. Criteria are proposed to compare and rank performance measures based on satellite characteristics, flood processes and computation time. As explained above, owing to the difficulty of modeling and simulating the various sources of spatially distributed observation errors, we resort to a “metaverification” approach (Murphy, 1996). Traditional performance measures based on pixel-to-pixel comparison are assessed, as are geometric performance measures based on spatial patterns and structure identification. For illustration, a flood event in the Garonne River in France was chosen, as it has been studied several times for flood applications in the context of parameter identification (Besnard & Goutal, 2011; El Garroussi et al., 2019; Nguyen et al., 2022).

The study discusses how to select performance measures to compare numerical simulations of a flood event versus a satellite image. The article is structured as follows. Section 2 describes the proposed methodology for selecting performance measures for comparing observed and simulated flood maps. Section 3 presents the materials for the Garonne case study: the flood model to be calibrated, the satellite image, and the tested performance measures. Section 4 reports the results of this numerical illustration. Section 5 gives an example of calibration, and the methodology is discussed. Finally, Section 6 draws a conclusion and presents future perspectives.

2. Methodology for Selecting Performance Measures

In the following, $O, S \in \{0, 1\}^{N_x \times N_y}$ refers to flood map matrices from observation and simulation respectively. The state of each pixel is dichotomized as dry (0) or flooded (1). The term $\mathcal{L}(\dots)$ denotes any performance measure. The flood model generates simulations based on parameter realizations ω_m , where $m \in [1; M]$ with M being the number of samples. $H(\omega_m)$ or $S(\omega_m)$ denote the simulated water depths or flood maps, respectively, of one member of the random sample. The number of simulations M is adjusted to ensure convergence of the statistical moments of the probability distribution of $\mathcal{L}(O, S(\omega_m))$. The coordinates of the pixel centers are stored in two matrices $X, Y \in \mathbb{R}^{N_x \times N_y}$. The total number of pixels $N_x \times N_y$ is denoted by N .

Our goal was to select the most suitable performance measure to evaluate flood event simulations versus satellite observations (see Figure 1, Section 2.1). The first selection step evaluates correlations between performance measures for an observed flood map (reference) and a sample of simulated flood maps. Such correlations can help reduce the number of performance measures for analysis (see Figure 1, Section 2.2). In this study, we develop a “metaverification” process, as introduced by Murphy (1996). Metaverification aims to identify performance measures that meet specific criteria, such as robustness with respect to translation/noise, etc. Depending on the nature of the problem (floods, temperature field estimation, rainfall analysis, etc.), these criteria may differ (Jolliffe & Stephenson, 2012), and here we propose specific criteria to compare flood maps based on simulations and on satellite observations (see Figure 1, Section 2.3). The evaluation criteria aim to select a performance measure that allows calibration within a random sample of simulations: that is, that, via a calibration procedure, can identify a parameter field despite limited observations and computational limitations. The criteria are not ordered by importance. The first criterion quantifies the performance measures' sensitivity to errors in flood magnitude: that is, spatial extent (see Figure 1, Section 2.3.1). Then, two criteria are used to evaluate sensitivity to displacement and noise errors. For these criteria, we consider a reference flood map (either simulated or observed) and generate variations in the map to reproduce errors in the observation or simulation (see Figure 1, Sections 2.3.2 and 2.3.3). The sensitivity of the performance measure to the variations in the flood map is evaluated numerically, relative to the unvarying reference map. Finally, a criterion is based on the computational requirements to estimate the performance measures (see Figure 1, Section 2.3.4). The first and second criteria were inspired by the Cloke and Pappenberger (2008) procedure applied to precipitation fields. In the present study, they are adapted for comparison of flood maps. Our selection methodology first analyzes each criterion independently. The performance measures that are not rejected by any of the criteria are then ranked (see Figure 1, Section 2.4), based on the possibility of identifying the same simulation as optimal with respect to observation errors.

In the following subsections, we present the data processing used to generate the reference observed flood map and the samples of simulated flood maps. Then, the correlation analysis, criteria and ranking workflow are described. Figure 1 shows all the steps of the methodology, with schematic examples to help practitioners understand and implement the methodology in any case study. The various steps in the Figure are described more exhaustively in the following subsections.

2.1. Data Processing

2.1.1. Observed Flood Maps O

An SAR observation is a grid of pixels of size $N_x \times N_y$, where each element corresponds to the backscatter intensity value (proportional to the energy received over the energy emitted by the satellite on the pixel), denoted as σ^0 , capturing the radar signal's interaction with the Earth's surface. Based on the backscattering intensity, it is possible to create a flood map O classifying each pixel as dry or flooded. This task can be carried out with various extraction procedures (Grimaldi et al., 2016; Landuyt et al., 2018; Schumann et al., 2023), but the present study relied on a single method, as this was not the focus of the study. The flood maps were extracted using a local thresholding procedure that dichotomized flooded and dry pixels based on the backscattering values; based on the histogram of a satellite observation, a threshold σ_{ref}^0 was determined. A more comprehensive description of the principle and algorithm can be found in Chini et al. (2017) and Appendix A. This method is widely used in flood mapping studies and is considered reliable for the extraction of flood extent (Hostache et al., 2018; Landuyt et al., 2018). The Joint Research Centre, for instance, uses this method to map floods for the Global Flood Monitoring product.

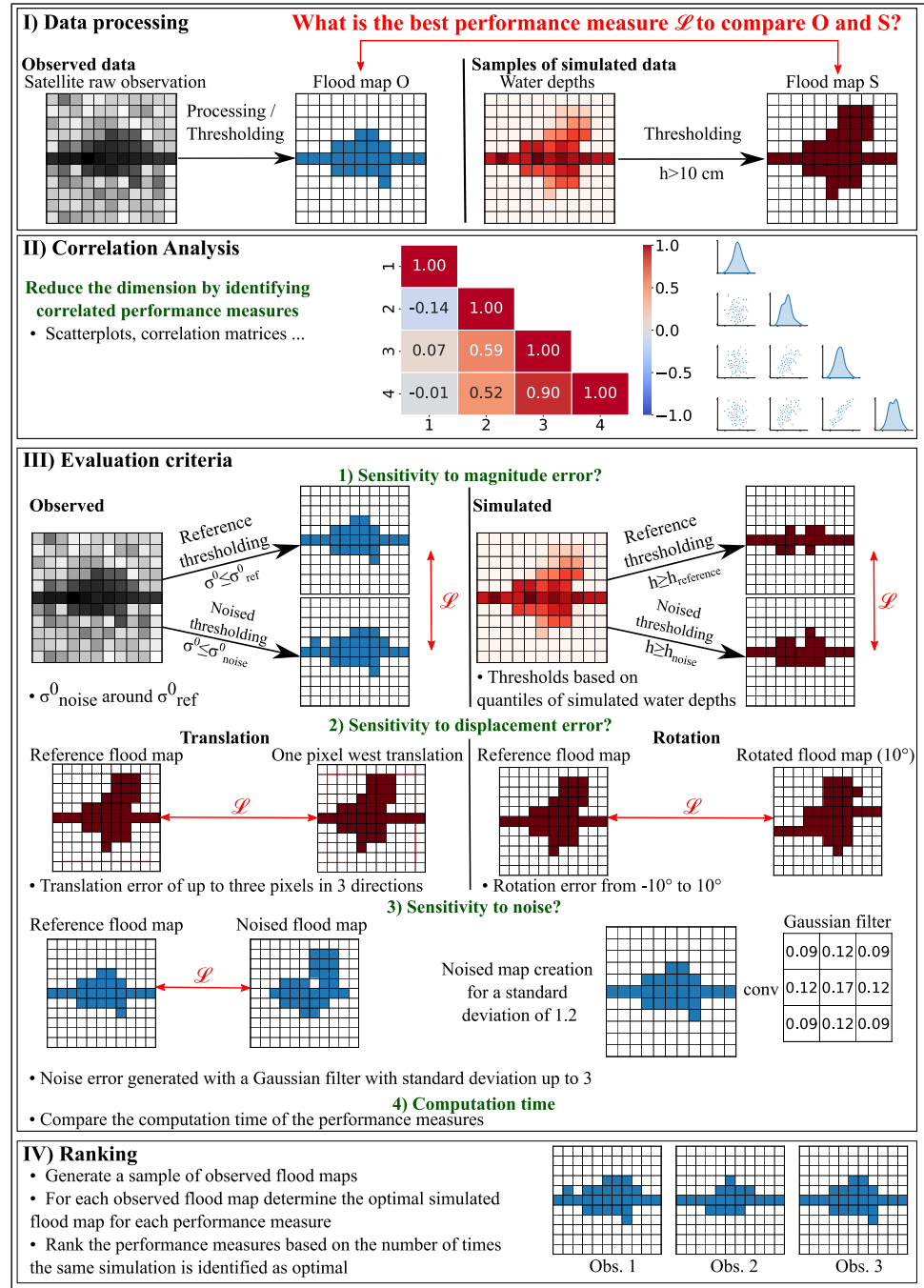


Figure 1. Step-by-step methodology to guide selection of a performance measure ($\mathcal{L}$) to compare observed (O) and simulated (S) flood maps.

2.1.2. Simulated Flood Maps S

Water depth is the value of interest and output of the simulation. However, to be compared with flood map observations, processing is needed. For each simulation, we defined a matrix $H \in \mathbb{R}^{N_t \times N_y}$ such that $H_{i,j}$ equals the

value of the finite element water depth at the center of the pixel (i, j) . This was then converted into a flood map matrix using a threshold τ water depth. S^τ denotes a map such that $S^\tau_{i,j} = \mathbb{1}_{H_{i,j} > \tau}$ for each pixel (i, j) of the observation. Here, we follow Aronica et al. (2002) and set τ at 10 cm; then, for the sake of clarity, we simply rewrite S^τ as S .

2.2. Correlation Analysis

The number of performance measures to be evaluated can be reduced by correlation analysis to eliminate redundancies and ensure that all performance measures identify distinct aspects of the errors. Correlations can be identified graphically on scatter plots, correlation matrix heatmaps, and principal component analysis (PCA) correlation circles. Practitioners can use any of these to identify linear correlations between performance measures. In the present study, all these techniques were used, and especially scatter plots. Scatter plots show the relationship between two variables. The alignment of the points in the scatter plots reflects the strength of the relationship. Scatter plots can identify linear and non-linear relationships between the performance measures, and linearly correlated performance measures can be excluded from further analysis.

2.3. Evaluation Criteria

2.3.1. Criterion 1: Sensitivity to Magnitude Error

A good performance measure should be significantly sensitive to changes in the magnitude of the quantity of interest: that is, the extent of the flood. If there are fewer or more flooded pixels in the domain, typically because of the sensitivity of the “flooded” status of a pixel (link to the minimum threshold τ in simulations, or to the post-processing of observations), the performance measure should be affected.

To quantify this sensitivity to variations (“errors”) in the magnitude of flood extent, following Cloke and Pappenberger (2008), we generated quantiles within each simulated or observed flood map after varying the main processing parameter in each case (the threshold τ or σ_{ref}^0). The relative variations can be analyzed for each performance measure as a function of quantiles, possibly after a preliminary statistical analysis of the relative variations in the samples, such as simulated flood maps resulting from a random distribution of the friction parameter or observed flood maps from different hyperparameterizations of the extraction algorithms. We conducted this analysis on simulations and observations independently.

2.3.1.1. Simulated Flood Maps

The volume of water in the simulated domain underlies the variations in flood extent. By choosing different water depth thresholds τ , we generated variations in flood extent to evaluate the sensitivity of the performance measure to magnitude errors. For each simulation output $H(\omega_m)$, flood maps were created with varying water depth threshold $\tau = h_l$ with $l \in \{10; 20; \dots; 80; 90\}$. The levels h_l were defined as empirical quantiles of the distribution of values $H_{i,j}$ with $i, j \in [1, \dots, N_x] \times [1, \dots, N_y]$. For every simulation, $H_{i,j}(\omega_m)$ can be ordered by increasing values $((H_k)_{(i,j)} | k = 1, \dots, N_x \times N_y)$, then $h_l = H_k$ with $k = \frac{l}{100} \cdot N_x \cdot N_y$.

Then, to check the behavior of a performance measure $\mathcal{L}$ to changes in flood extent, the map with the quantile h_{50} was chosen as a reference and compared with the other quantiles. $\mathcal{L}(S^{h_{50}}(\omega_m), S^{h_l}(\omega_m))$ was computed for all simulations $1 \leq m \leq M$ and all quantiles. For each quantile, the numerical evaluation of the performance measures was visualized on box-plots representing the range of values for the M simulations.

2.3.1.2. Observed Flood Maps

The observed flood maps may vary depending on the extraction procedure parameters. In this study, identification of σ_{ref}^0 was subject to uncertainties due to measurement and post-processing errors. To evaluate the behavior of the performance measures with these uncertainties, flood maps with other thresholds around σ_{ref}^0 were created. By adjusting the threshold, flood maps with varying degrees of flooding were generated (similarly to the magnitude errors for the simulated fields). The impact of threshold error is illustrated in Figure 2. Ten flood maps, denoted as $(O^{\sigma_i^0})_{-5 \leq i \leq 5}$, with threshold values $\sigma_i^0 \in \{0.95 \sigma_{ref}^0, \dots, 0.99 \sigma_{ref}^0, \sigma_{ref}^0, 1.01 \sigma_{ref}^0, \dots, 1.05 \sigma_{ref}^0\}$, were created and compared with the reference flood map. The flood maps with the lowest errors with respect to the reference can be

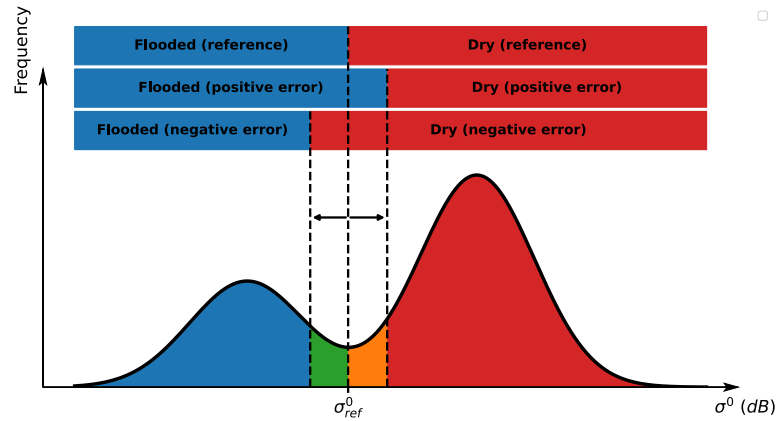


Figure 2. Illustration of the creation of observed flood maps with errors (positive for $\sigma^0 \geq \sigma_{ref}^0$, and negative for $\sigma^0 \leq \sigma_{ref}^0$) in threshold identification on bimodal Gaussian distributions of σ^0 .

expected to have performance measure values closer to the reference, similarly to the previous test on the simulated fields.

2.3.2. Criterion 2: Sensitivity to Displacement Error

Displacement errors occur when extracting data from satellite images (e.g., georeferencing inaccuracies, map projection, platform orientation, and so on.). Rotation and translation errors should be corrected to avoid location errors, and a performance measure should be sensitive to these types of error.

2.3.2.1. Translation Error

SAR observations can be subject to spatial translation errors, and a good performance measure should be significantly sensitive to such errors. To quantify this sensitivity, we generated translation errors on flood maps in various directions. Each flood map $S(\omega_m)$ was translated by 1–3 pixels in three directions: 1 north (1N), 1 west (1W), 1 northwest (1NW), 2 north (2N), 2 west (2W), 2 northwest (2NW), 3 north (3N), 3 west (3W), and 3 northwest (3NW). Longer distances were not considered, as they do not represent probable translations from geolocation errors or the extraction of flood maps (Schumann et al., 2008). For a grid of pixels with a resolution of 10 m, center-to-center distances between the reference maps and the translated maps ranged from 10 to 42 m (three pixels in the northwest direction). The translated fields are denoted as $S'(\omega_m)$ with $t \in \{0; 1N; 1W; 1NW; 2N; 2W; 2NW; 3N; 3W; 3NW\}$, and $S^0(\omega_m) = S(\omega_m)$. The value $\mathcal{L}(S(\omega_m), S'(\omega_m))$ was computed for all simulations $1 \leq m \leq M$ and all translations. The evaluation of the performance measure for a translated field against the reference field should deviate from the optimal value of the performance measure with increasing distance.

2.3.2.2. Rotation Error

The observation may also contain rotational inaccuracies due to small changes in the sight angle and distortion errors (Sansosti et al., 2006). The simulated flood maps were rotated, counterclockwise around the center, by different angles and compared to the reference flood maps. The coordinates of the rotated field were calculated as follows:

$$\begin{bmatrix} (X_{rotated})_{i,j} \\ (Y_{rotated})_{i,j} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} X_{i,j} \\ Y_{i,j} \end{bmatrix}, \quad (1)$$

where θ is the angle of rotation. The simulated flood maps were rotated arbitrarily by $\theta \in \{-10^\circ, -5^\circ, -2^\circ, -1^\circ, 0^\circ, 1^\circ, 2^\circ, 5^\circ, 10^\circ\}$ and the rotated fields are denoted as $S^\theta(\omega_m)$ with $S^0(\omega_m) = S(\omega_m)$. Then, similarly to the translation displacement, $\mathcal{L}$ was calculated for each configuration against the reference simulated flood maps.

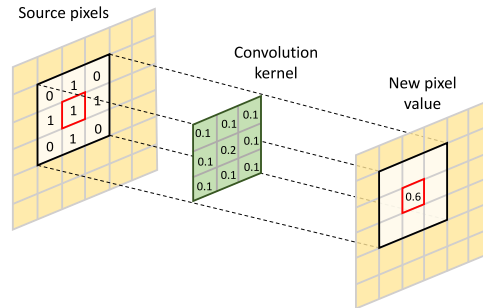


Figure 3. Example of matrix convolution by a convolution kernel.

The evaluation of the performance measure for a rotated field should deviate from the optimal value with increasing angles.

2.3.3. Criterion 3: Sensitivity to Noise

A good performance measure should naturally be sensitive to the noise spatially distributed within the satellite observation images, before or after processing (here, as an extracted flood map). The SAR backscatter values are a function of speckle (grainy noise in SAR images from wave interference, affecting pixel intensity) and of the underlying land use, which often exhibits a spatial autocorrelation, such as flooded areas (Ulaby et al., 1986). The flooded pixels can be subject to spatial autocorrelation over hundreds of meters (Waller et al., 2018), with neighboring pixels more likely to share the same state.

To evaluate this sensitivity, Cloke and Pappenberger (2008) generated spatial autocorrelation using Sequential Gaussian Simulations (SGS) with various spatial correlations based on a variogram computed on the observation. However, SGS is too computationally expensive for large numbers of pixels (here, more than 6,113,000) because of the variogram modeling and kriging steps. Another possibility is to introduce noise with a Gaussian filter. Gaussian filtering aims to smooth an image by replacing each pixel value with a linear combination of its neighbors. Practically, it involves performing a convolution between an input matrix and a convolution kernel. An illustration is given in Figure 3, with a kernel of size 3.

The reference observed flood map was noised for different kernel sizes (3, 5, 7, and 9) with $s \in \{0.1, 0.2, \dots, 2.9, 3.0\}$ where s is the standard deviation. After convolution, pixels with values above 0.5 were considered flooded, and otherwise dry. The noised observations denoted as O^s were compared to the reference computing $\mathcal{L}(O, O^s)$. For a Gaussian filter, the kernel is computed as follows:

$$G(X_{i,j}, Y_{i,j}) = \frac{1}{2\pi s^2} e^{-\frac{X_{i,j}^2 + Y_{i,j}^2}{2s^2}}. \quad (2)$$

The kernel matrix is then normalized so that the sum equals one.

2.3.4. Criterion 4: Computation Time

The cost of computing a performance measure could be prohibitive; depending on the computational capability available, the user may be interested in more cost-effective implementations. For each call to $\mathcal{L}$, computation time is evaluated. Preprocessing of images, such as projections or flood mapping, is not taken into account for the computation time. Selection of a performance measure may depend on the efficiency of implementing the performance measures by the developers.

2.4. Ranking Performance Measures

After the selection phase, the practitioner may still have several performance measures available, and the last step is to rank them. The proposed ranking process is based on the sensitivity of parameter identification to observation errors. A sample of 100 flood maps (convergence of the ranking on our test case) with different processing (e.g.,

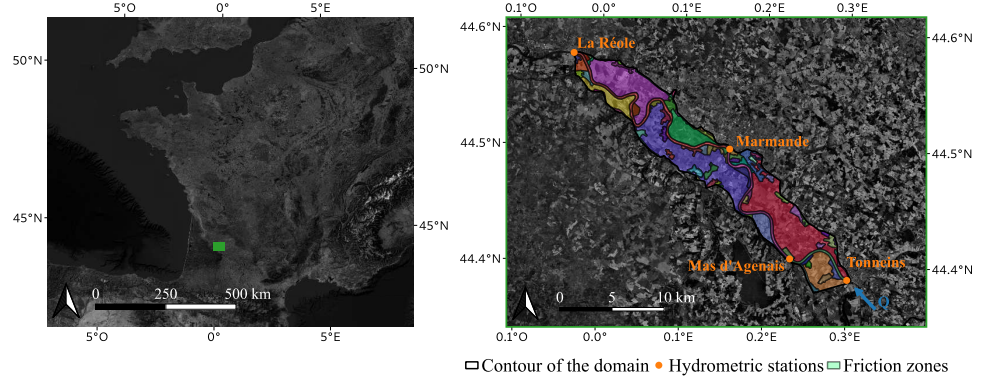


Figure 4. Visualization of the study area on the Garonne River in France. The dots represent in-situ gauging stations and the colored subdomains indicate the floodplain friction zones.

different thresholds distinguishing dry and flooded pixels) following the same process was created, as in Section 2.3.1, with uniformly distributed thresholds $(\sigma_j^0)_{1 \leq j \leq 100}$ between minimum and maximum values. The choice of values was based on a proposal by Hostache et al. (2009) whereby pixels that are uncertain as to flooded or dry status in the yellow and red zones of Figure 2 are constrained by two thresholds: σ_1^0 is the minimum value of σ^0 in grassland regions of the dry area in floodplains, and σ_{100}^0 is the maximum value of σ^0 in permanent water bodies. For each observation O_j^0 with $1 \leq j \leq 100$, the simulated flood map minimizing performance measure discrepancy is determined as:

$$E_j = \min_{1 \leq m \leq M} |\mathcal{L}(O_j^0, S(\omega_m)) - \mathcal{L}(O_j^0, O_j^0)|. \quad (3)$$

$E = \{E_j | 1 \leq j \leq 100\}$ describes the set of simulations determined in Equation 3. For each performance measure, the maximum number of times a simulation is identified in set E is computed as follows:

$$n_{\max} = \max_{1 \leq j \leq 100} |\{E_{j'} | E_{j'} = E_j\}| \quad (4)$$

where $\{E_{j'} | E_{j'} = E_j\}$ denotes the subset of simulations that are equal to E_j . The performance measures are ranked according to the $n_{\max}$ values. The performance measures that are most robust with respect to observation errors are those in which the same simulation is identified the most often in E .

3. Materials for Numerical Illustration of the Methodology

The methodology was illustrated by a numerical application to a specific case of a flood that occurred in France, for which satellite images were freely available (see Section 3.1). Numerical simulations were performed using a standard flood model (see Section 3.2). The performance measures used to compare the observed and simulated flood maps are presented in Section 3.3.

The study area was the Garonne River between Tonneins and La Réole, in southwest of France (see Figure 4). The reach is about 50 km long, with a mean width of 250 m under normal flow conditions. The floodplain is 1–4 km wide, and mainly comprises rural areas used for agricultural purposes, with the remainder consisting of small patches of vegetated and urban areas. Since the nineteenth century, following a major flood in 1875, the floodplains have been equipped with dikes and levees to protect urban areas and infrastructure from flooding. This area has already been used numerous times for the study of numerical models of floods, and we built on these previous efforts in constructing the numerical model (see Section 3.2).

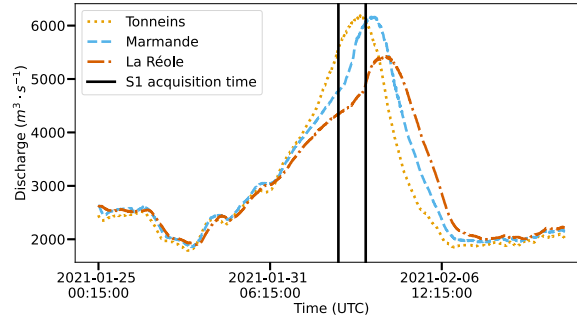


Figure 5. Measured discharge at the gauging stations (Tonneins, Marmande, and La Réole) and Sentinel-1 (S1) acquisition times during the February 2021 flood event.

3.1. Satellite Observation

In February 2021, a significant flood occurred in the study area. The flood event was observed twice by the Sentinel-1 C-band SAR instrument at 5.405 GHz. Acquisition times are reported in Figure 5, along with the discharge measured every 15 min at three in-situ stations. All times reported below are given in Coordinated Universal Time (UTC).

The present study used the first satellite observation, on February 2, at 5:55 p.m. The image was downloaded from ASF Data Search Vertex (<https://search.asf.alaska.edu/>). The observation extracted at “LEVEL-1 GRD” with VH polarization is a preprocessed $N_x \times N_y$ grid of pixels (involving orbit correction, radiometric calibration, noise reduction, thermal noise removal, terrain correction, and georeferencing) at a resolution of 10×10 m. Here, the dimensions of the image are $N_x = 2644$ and $N_y = 2312$. Then, each pixel is classified as flooded or dry, to create a flood map as described in Section 2.1.

3.2. Numerical Simulation by a Hydrodynamic Model

To simulate flood events, a numerical solution of the 2D shallow-water equations (SWEs) (5–7) was used. The SWEs in a Cartesian coordinate system where gravity acts uniformly as $-g\vec{e}_z$ can be written as:

$$\frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} = 0, \quad (5)$$

$$\frac{\partial(hu)}{\partial t} + \frac{\partial(hu^2)}{\partial x} + \frac{\partial(huv)}{\partial y} = -gh \frac{\partial \eta}{\partial x} + \nabla \cdot (h \nu_e \nabla u) - \frac{1}{\rho} \tau_{bx}, \quad (6)$$

$$\frac{\partial(hv)}{\partial t} + \frac{\partial(hv^2)}{\partial y} + \frac{\partial(huv)}{\partial x} = -gh \frac{\partial \eta}{\partial y} + \nabla \cdot (h \nu_e \nabla v) - \frac{1}{\rho} \tau_{by}. \quad (7)$$

These are standardly used to model free-surface flows with water depth $h \geq 0$ and a depth-averaged velocity $u\vec{e}_x + v\vec{e}_y$ as unknown variables that are functions of x , y and time $t \in [0, T)$. The water surface is denoted $\eta = h + z_b$, where $z_b(x, y)$ is the given bottom elevation, $\nu_e > 0$ a constant viscosity, and $\tau_{bx}\vec{e}_x + \tau_{by}\vec{e}_y$ the bed shear stress function of h , u , and v . The Manning-Strickler formula (Manning, 1891) is used for the bed shear stress:

$$\begin{cases} \tau_{bx} = \frac{-g \cdot u}{h^{4/3} \cdot K_s^2} \sqrt{u^2 + v^2} \\ \tau_{by} = \frac{-g \cdot v}{h^{4/3} \cdot K_s^2} \sqrt{u^2 + v^2} \end{cases}, \quad (8)$$

where, for application to real floods, the spatially distributed Strickler coefficient K_s needs calibrating, as a friction parameter that is a function of x , and y (Morvan et al., 2008).

Table 1
Range of the Friction Values for Each Floodplain Land Use

Land use	Range ($\text{m}^{1/3} \cdot \text{s}^{-1}$)
Waterbodies	$\mathcal{U}(32.5, 37.5)$
Fields and meadows without crops	$\mathcal{U}(17.5, 22.5)$
Cultivated fields with low vegetation	$\mathcal{U}(15, 20)$
Cultivated fields with high vegetation	$\mathcal{U}(10, 15)$
Shrublands and undergrowth areas	$\mathcal{U}(8, 12)$
Areas of low urbanization	$\mathcal{U}(8, 10)$
Highly urbanized areas	$\mathcal{U}(5, 8)$

TELEMAC-2D (T2D), a module of the open-source hydro-informatics system openTELEMAC (www.open-telemac.org) (Hervouet, 2007), was used to solve Equations 5–7, complemented by boundary conditions. It relies on a continuous piecewise-linear finite element approach, based on an unstructured triangular mesh. The N edge-by-edge scheme and the Positive Streamwise Invariant distributive scheme were used to compute the advection of velocity and water depth, respectively. An important feature for floodplain modeling is the wetting and drying of grid elements, taken into account here by correction of the free surface gradient (Hervouet, 2007).

For the present study, the T2D model used a simplicial mesh made up of $\mathcal{N} = 41,416$ nodes, and the size of the edges of the grid varied from 40 m in the channel to 80 m around the dikes and 150 m in the floodplains (Figure 4). To accurately represent dikes, the mesh was constrained along different lines. Channel bathymetry was reconstructed with bottom elevations measured from 70 cross-sections. Floodplain topography was determined from IGN (French National Institute of Geographic and Forest Information) maps and aerial photographs. A more complete description of the construction of the model can be found in Besnard and Goutal (2011). The model was initialized at $t = 0$ with a previous steady-state simulation with upstream discharge of $2,300 \text{ m}^3 \cdot \text{s}^{-1}$. For the boundary conditions, a time-varying hydrograph was imposed upstream at Tonneins (see Figure 5 for the flood event under study) and a rating curve linking water depth to discharge was imposed downstream at La Réole.

The friction parameter K_s was piecewise constant over the simulation domain: that is, with a uniform value over subdomains. The riverbed was divided into 3 subdomains, based on prior calibration of the model on historical floods: $K_s = 45 \text{ m}^{1/3} \cdot \text{s}^{-1}$ between Tonneins and Mas d'Agenais, $K_s = 38 \text{ m}^{1/3} \cdot \text{s}^{-1}$ between Mas d'Agenais and Marmande, and $K_s = 40 \text{ m}^{1/3} \cdot \text{s}^{-1}$ between Marmande and La Réole. The floodplain was divided into 92 subdomains, corresponding to areas of homogeneous land use based on the 2018 Corine Land Cover data set (<https://land.copernicus.eu/en/products/corine-land-cover/clc2018>), as visualized in Figure 4. For each land use class, the prior range of Strickler values was defined as shown in Table 1. The range limits were fixed according to expert knowledge, primarily based on 1D considerations and laboratory experiments (Barnes, 1967; Chow et al., 1988). A uniform distribution was considered since the shape of the Strickler distribution for each land use is not described in the literature.

3.3. Performance Measures

The following subsections detail the performance measures used in this study. Several are already known to the hydraulic community, mostly based on pixel-to-pixel comparisons of 2D flood maps (Aronica et al., 2002; Grimaldi et al., 2016; Pappenberger et al., 2007). Geometric performance measures based on shape analysis or texture identification are also presented. Twenty-four are evaluated here, chosen from the literature based on ease of implementation in the context of the present case study (Wilks, 2011).

3.3.1. Pixel-To-Pixel Comparison of Flood Maps

Pixel-to-pixel performance measures compare the number of pixels that match or do not match between a simulation and an observation. There are four possible outcomes, which can be represented in a 2×2 confusion matrix (or contingency table), as shown in Figure 6. At the grid points, a hit occurs when a pixel is correctly predicted to be flooded (True Positive (TP)) and a correct rejection (True Negative (TN)) occurs when it is

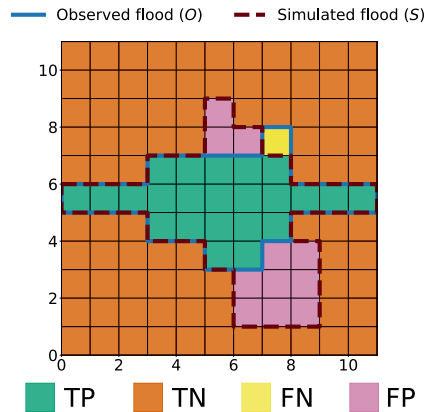


Figure 6. Classification of True Positive, True Negative, False Negative, and False Positive for the comparison of two schematic observed and simulated flood maps.

correctly predicted to be dry. A false alarm (False Positive (FP)) and a miss (False Negative (FN)) occur when a pixel is simulated as flooded or dry, respectively, whereas the opposite is observed.

Numerous performance measures combine TP , TN , FP , and FN ; the most widely used are listed in Table 2. The Table reports the range of possible values, the perfect value (for the same flood maps), and a description of each performance measure. In the present study, 15 performance measures were evaluated. F_β was calculated for 3 values, with $\beta \in \{1; 1.5; 2\}$. Many variants have been formulated in the literature to reflect the need to assess classification from different perspectives and accommodate the specific requirements of different domains. For example, for floods in sensitive areas, minimizing FN (ensuring high recall) might be more important for decision-makers than ensuring high precision by minimizing FP . Other performance measures, such as the Matthews correlation coefficient (MCC) or Cohen's Kappa formula (κ), aim at balancing the variables of the confusion matrix to avoid bias from unbalanced classes. Although FPR , TPR , PPV , and NPV are naive performance measures, we used them to assess the rejection capability of the proposed methodology as part of our testing process.

Table 2

Most Widely Used Performance Measures for Pixel-To-Pixel Comparison Based on the Confusion Matrix Adapted From Grimaldi et al. (2016)

Name	Evaluation	[Min, max, perfect value] description
False positive rate (FPR)/Precision (F)	$FP/(FP + TN)$	[0, 1, 0] Proportion of overprediction of flooded areas
True positive rate (TPR)/Recall (H)	$TP/(TP + FN)$	[0, 1, 1] Proportion correct of observed flooded areas
Positive predictive value (PPV)	$TP/(TP + FP)$	[0, 1, 1] Proportion of positive pixels that are true positive
Negative predictive value (NPV)	$TN/(FN + TN)$	[0, 1, 1] Proportion of negative pixels that are true negative
Peirce skill score (PSS)	H-F	[-1, 1, 1] Maximizing the difference between H and F
Accuracy (ACC)	$(TP + TN)/(TP + FP + TN + FN)$	[0, 1, 1] Proportion correct of total domain area
F_β -score	$(1 + \beta^2) \cdot (H \cdot F) / (\beta^2 \cdot F + H)$	[0, 1, 1] Combine precision and recall
Matthews correlation coefficient (MCC)	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$	[-1, 1, 1] Balance of the four categories
Bias (B)	$(TP + FP)/(TP + FN)$	[0, ∞ , 1] Ratio highlighting over/underprediction
Critical success index (CSI)	$TP/(TP + FP + FN)$	[0, 1, 1] Adaptation of ACC to focus on flooded areas
$F^{<3>}$	$(TP - FN)/(TP + FP + FN)$	[-1, 1, 1] Designed to penalize underprediction of flooded areas
$F^{<4>}$ /Measure of fit	$(TP - FP)/(TP + FP + FN)$	[-1, 1, 1] Designed to penalize overprediction of flooded areas
Cohen's Kappa formula (κ)	$\frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)}$	[-1, 1, 1] Balance of the four categories

Table 3
List of Evaluated Geometric Performance Measures

Name	Method type	[Min, max, perfect value] description
Hausdorff distance (d_H)	Distance based	$[0, \infty, 0]$ Maximum distance between points of two sets
Modified Hausdorff distance (d_{MH})	Distance based	$[0, \infty, 0]$ Mean of the distances from each point in one set to the closest point in the other set
Procrustes analysis (P^2)	Shape alignment	$[0, \infty, 0]$ Euclidean distance between shapes after rotation, scaling, and rotation
Sliced-Wasserstein distance (W_2)	Distribution based	$[0, \infty, 0]$ Distance between probability distributions
Normalized Mutual Information (NMI)	Information theory	$[1, 2, 2]$ Quantifies similarity using entropy
Flood Skill Score (FSS_n)	Neighborhood	$[0, 1, 1]$ Quantifies overlap by comparing the flooded fraction over multiple neighborhood scales

3.3.2. Geometric Comparison of Flood Maps

Another possibility for comparing 2D flood maps is to examine their geometric properties. These performance measures may provide more information, since floods have a spatial structure. In this subsection, each performance measure is briefly introduced, and further information can be found in the cited references. The performance measures are summarized in Table 3 with a brief description, the range of possible values, and the perfect value for each performance measure.

3.3.2.1. Distance Based Measures

The performance measures introduced below are based on computing distances on set of points (point-to-point or between contours).

3.3.2.1.1. Hausdorff Distance

The Hausdorff distance (Hausdorff, 1927) measures the greatest distance between any point in one set (here, the observed flooded pixels) and the closest point in another set (here, the simulated flooded pixels). As an illustration, in Figure 7, the green line shows the greatest distance from any point in set O to the closest point in set S , and the orange line shows the maximum distance from any point in set S to the closest point in set O . The length of the orange line is the Hausdorff distance. The Hausdorff distance denoted as d_H is formulated as follows:

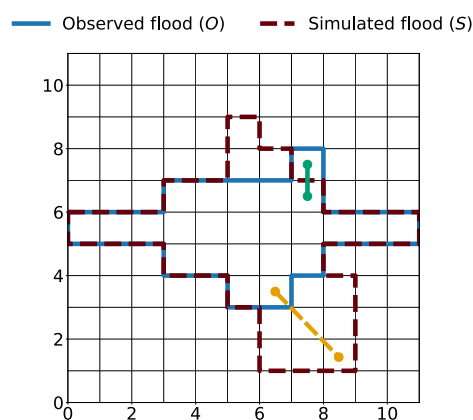


Figure 7. Visualization of the greatest minimum distance from the observed flood (O) to the simulated flood (S) in green and from S to O in dotted orange.

$$d_H(O, S) = \max \left\{ \max_{(i,j), S_{ij}=1} \left\{ \min_{(k,l), O_{kl}=1} \delta((i,j), (k,l)) \right\}, \max_{(i,j), O_{ij}=1} \left\{ \min_{(k,l), S_{kl}=1} \delta((i,j), (k,l)) \right\} \right\}, \quad (9)$$

where

$$\delta((i,j), (k,l)) = \sqrt{(X_{ij} - X_{kl})^2 + (Y_{ij} - Y_{kl})^2}. \quad (10)$$

3.3.2.1.2. Modified Hausdorff Distance

Hausdorff distance can be highly sensitive to outliers. To address this issue, the Modified Hausdorff distance was introduced by Dubuisson and Jain (1994). Instead of taking the greatest minimum distance between the sets, it averages the distances between the closest points in both sets. The Modified Hausdorff distance is denoted as d_{MH} and formulated as:

$$d_{MH}(O, S) = \max \left\{ \frac{1}{|O|} \sum_{(i,j), O_{ij}=1} \min_{(k,l), S_{kl}=1} \delta((i,j), (k,l)), \frac{1}{|S|} \sum_{(k,l), S_{kl}=1} \min_{(i,j), O_{ij}=1} \delta((i,j), (k,l)) \right\}, \quad (11)$$

where $|O| = \sum_{i,j} |O_{ij}|$ and $|S| = \sum_{i,j} |S_{ij}|$.

3.3.2.2. Other Performance Measures

The following performance measures are grouped together, as they are not directly related to computing the distances between individual pixels, but are related to the cost of matching whole sets of flooded pixels.

3.3.2.2.1. Procrustes Analysis

Procrustes analysis (Gower, 1975) is a shape alignment method, comparing two shapes by removing differences in position, size and orientation. It aims to find an optimal transformation (scaling, translation and rotation) between O and S , to minimize the sum of square differences between the transformed set of points, as illustrated in Figure 8.

Analysis begins by standardizing O and S into two matrices $\mathcal{M}_1$ and $\mathcal{M}_2$, centered around their origins and $\text{Tr}(\mathcal{M}_i \cdot (\mathcal{M}_i)^T) = 1$ with $i \in \{1; 2\}$. Then an optimal transformation represented by an orthogonal matrix Q and a scaling factor c is determined using singular value decomposition. The objective is to minimize the quantity P^2 calculated as:

$$P^2 = d_e(\mathcal{M}_1, c\mathcal{M}_2Q), \quad (12)$$

$$\text{with } d_e(O, S) = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (O_{ij} - S_{ij})^2}$$

3.3.2.2.2. Sliced-Wasserstein Distance

The Wasserstein distance, derived from Optimal Transport, measures the minimal cost of moving mass to transform one distribution into another, based on a defined cost of moving a unit of mass from one position to another. For instance, Figure 9, shows the transport plan (where and how much mass to move from one pixel to another) and the cost matrices for moving from an observed flood map to a simulated flood map.

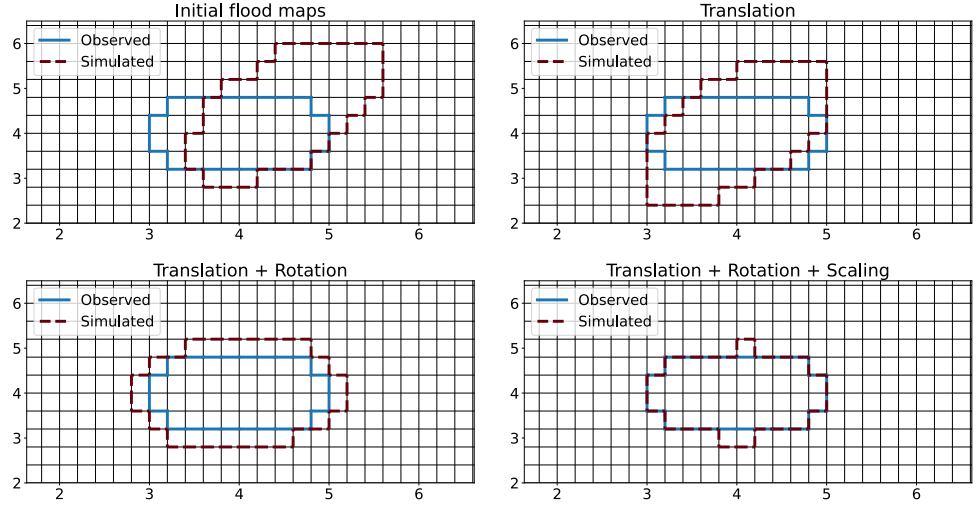


Figure 8. Optimal transformation between two schematic flood maps achieved by translating, rotating, and scaling the simulated flood map (brown contour).

To compute the Wasserstein distance on 2D flood maps, we consider $\text{vec}(O)_{\text{norm}}, \text{vec}(S)_{\text{norm}} \in [0, 1]^N$ the vectorization of O and S normalized such that $\|\text{vec}(O)_{\text{norm}}\|_1 = 1$ and $\|\text{vec}(S)_{\text{norm}}\|_1 = 1$ (e.g., blue and brown histograms in Figure 9). We define a matrix of cost $C \in \mathbb{R}^{N \times N}$, defined as $C_{ij} = \delta(i, j)$, representing the cost of moving mass from pixel i to j in O and S respectively; here, i and j refer to two pixel centers, while δ is the usual Euclidean distance in $\mathbb{R}^2$ as in Equation 10 as illustrated in the middle of Figure 9. Then, a possible performance measure is the 2-Wasserstein distance:

$$W_2(O, S) = \left(\min_{\gamma \in \Pi} \langle \gamma, C \rangle \right)^{1/2} = \left(\min_{\gamma \in \Pi} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} C_{ij} \right)^{1/2}. \quad (13)$$

with $\Pi = \{A \in [0, 1]^{N \times N} \mid \sum_{i=1}^N \sum_{j=1}^N A_{ij} = 1, A\mathbf{1} = \text{vec}(O)_{\text{norm}}, A^T\mathbf{1} = \text{vec}(S)_{\text{norm}}\}$ the optimal transport plan (on the right in Figure 9) and $\mathbf{1} \in \mathbb{R}^N$ the vector of ones. However, numerically computing Equation 13 is costly. To reduce the computational burden, we consider the Sliced-Wasserstein distance (Bonneel et al., 2015) as a practical alternative to Equation 13. It is a Monte Carlo approximation of W_2 obtained from K 1D projections chosen randomly. Each pixel of the 2D flood maps is projected onto a 1D line (a direction along the unit circle), making computation of the Wasserstein distance faster by reducing the dimensionality of the cost matrix from $N \times N$ to N . In this study, the number of projections arbitrarily fixed at $K = 1,000$.

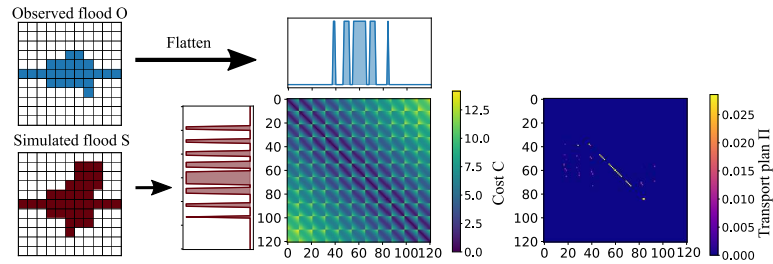


Figure 9. Visualization of the cost matrix and of the optimal transport plan to transform a simulated flood map into an observed flood map.

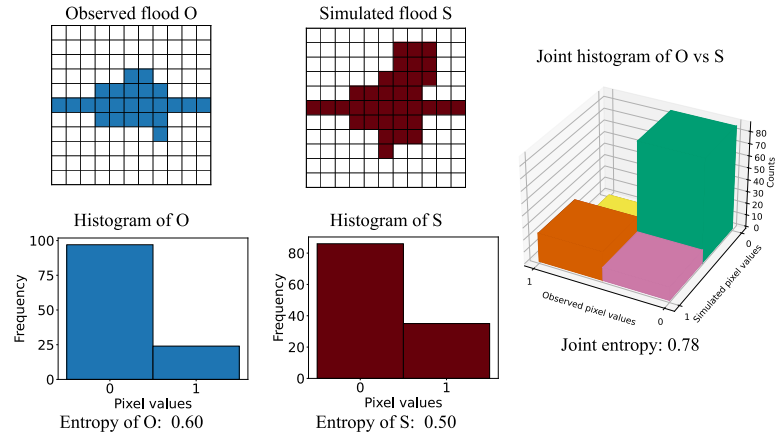


Figure 10. Histograms and calculated marginal and joint entropies for two (observed (O) and simulated (S)) schematic flood maps. In this case, the Normalized Mutual Information equals 1.41.

3.3.2.2.3. Normalized Mutual Information

Normalized Mutual Information (NMI) (Studholme et al., 1999) was developed for automating medical image alignment by measuring the content of information in images, using information theory. NMI is computed using the entropy (quantifying the amount of information) of the two flood map histograms and also their joint histograms. Figure 10 shows an illustration of these histograms and related entropies for a simple case. For instance, the green bar on the joint histogram represents the number of pixels classified as dry in both O and S . The more redundant information an image contains (e.g., uniform background), the lower its entropy.

NMI can be formulated as follows:

$$NMI(O, S) = \frac{H(O) + H(S)}{H(O, S)}, \quad (14)$$

where $H(O) = -\sum_{i=1}^c p_i \log(p_i)$ represents the Shannon entropy for matrix O , with c the number of classes (2 with flooded and dry pixels) and p_i the probability of class i .

$H(O, S) = -\sum_{i=1}^c \sum_{j=1}^c p_{ij} \log(p_{ij})$ with p_{ij} the joint probability between classes i in O and j in S .

3.3.2.2.4. Fraction Skill Score

The Fraction Skill Score (FSS) (Roberts & Lean, 2008) was introduced in meteorology to evaluate the performance of spatial forecasts. It compares the spatial patterns of a forecast and an observation, taking into account of the scale of the features. The FSS is based on a neighbor approach, in which the neighborhoods of simulated and observed pixels are compared. For every grid point, the number of flooded pixels within a window of size n around it over the total number of pixels in the window (n^2) is computed. From this process, 2D fraction fields are created for the observed and simulated flood maps called, respectively, O^n and S^n , as illustrated in Figure 11, with $n = 3$. The mean square error (MSE) for the observed and simulated fractions of a neighborhood of length n is given by:

$$MSE_n = \frac{1}{N_x \cdot N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{i,j}^n - S_{i,j}^n]^2 \quad (15)$$

The worst possible MSE between a simulation and an observation is calculated as:

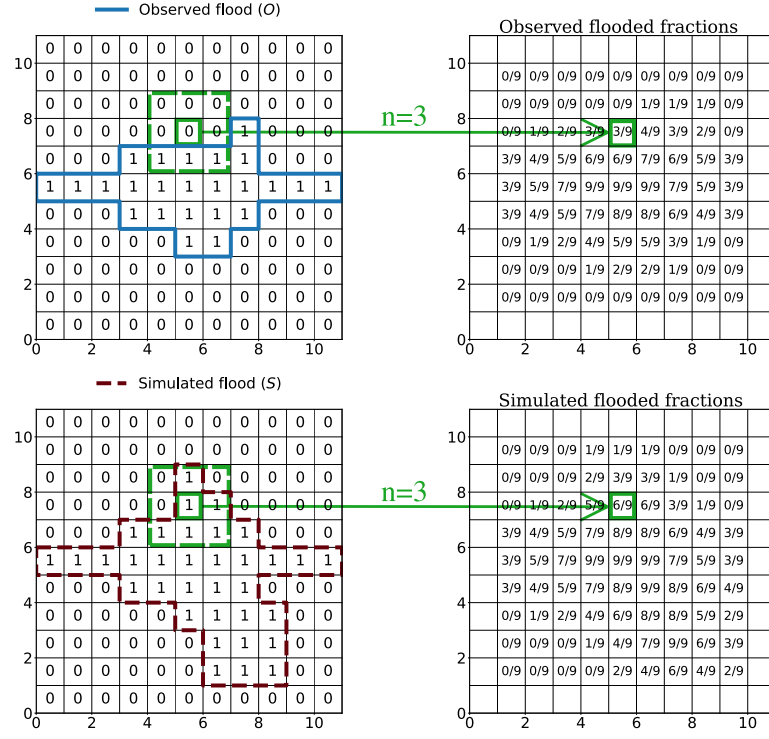


Figure 11. Visualization of the observed and simulated flooded fractions maps from schematic flood maps for a window of size 3×3 .

$$MSE_{n(ref)} = \frac{1}{N_x \cdot N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [(O_{i,j}^n)^2 + (S_{i,j}^n)^2] \quad (16)$$

Then, the FSS for each grid level is calculated as:

$$FSS_n = 1 - \frac{MSE_n}{MSE_{n(ref)}} \quad (17)$$

In the present study, FSS scores were computed for different levels $n \in \{1; 3; 5; 7\}$.

4. Numerical Illustration of the Methodology

The proposed methodology for selecting a performance measure was tested on the flood model and satellite observation described in Sections 3.1 and 3.2, respectively for the performance measures summarized in Tables 2 and 3. For the flood model, we considered a random sample of 4,000 simulations generated from 4,000 values of the 92-dimensional parameter field (the floodplain friction in Figure 4). The random parameter values were generated independently, using the distributions provided in Table 1. The number of $M = 4,000$ model evaluations was considered satisfactory to obtain converged statistics for all performance measures (see Appendix B). Each simulation cost 18 hr to compute $T = 13$ days of physical time on an Intel(R) Xeon(R) Platinum 8260 CPU at 2.40 GHz.

4.1. Correlation Analysis

First, the correlations between the performance measures were evaluated by computing $\mathcal{L}(O, S(\omega_m))$. The scatter plots in Figure 12 provide a comprehensive view of the pairwise relationships between the geometric performance

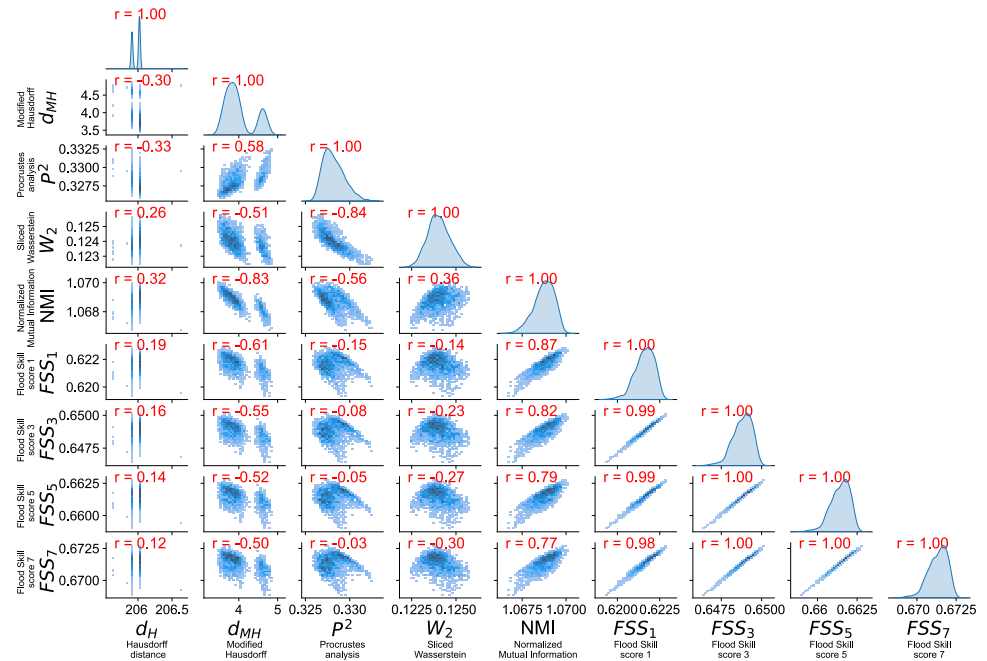


Figure 12. Scatter plots for geometric performance measures when evaluating the reference flood map with 4,000 simulated flood maps, with r being the Pearson correlation coefficient.

measures; scatter plots for the pixel-to-pixel performance measures are presented as Supporting Information S1. On the diagonal, the probability distribution of each performance measure was plotted, and the pairwise relationships between performance measures are off the diagonal. The linear correlations between each pair of measures were evaluated on Pearson correlation coefficient (from -1 for a perfect negative linear relationship to 1 for a perfect positive relationship, and 0 for no linear relationship), written in red in the Figures for each pair of performance measures. Linear relations existed between some performance measures, such as NPV , $F_{\beta=2}$, and $F^{<3>}$, or between PPV , $F^{<4>}$, and ACC , for the pixel-to-pixel performance measures. For the geometric performance measures in Figure 12, the FSS values for different window sizes are correlated almost linearly. Then, from these linear correlations, the number of performance measures in the analysis could be reduced, since, when performance measures correlate linearly, one measure can be deduced from the other. However, in this study, the correlations were not used directly, and all the performance measures were evaluated for the various criteria. Although, the performance measures may correlate in this data set (comparing O and $S(\omega_m)$), this does not guarantee that the correlation will hold for evaluation of the performance measure with variations in the flood maps (shifted, translated, and so forth.). For instance, in Section 4.4, PPV is not sensitive to noise errors, whereas $F^{<4>}$ is, despite the fact that they correlate for parameter variations. Additional analysis of the correlations is discussed in Appendix C. We can also note, on the diagonal of the scatter plots, that the probability distributions of the performance measures are unimodal. Consequently, optimal parameter values exist: a single value for most performance measures, and 2 values for d_H and d_{MH} (with a two-peak “bimodal” distribution).

4.2. Criterion 1: Sensitivity to Magnitude Errors

4.2.1. Simulated Flood Maps

Water depth distributions were computed for the 4,000 simulations, and we generated a simulated flood map thresholded for the various quantiles of the distributions. An example of flood map generation with two thresholds for one of the simulations can be found as Supporting Information S1. The results of the magnitude errors experiment on the simulated flood maps are shown in Figure 13, with red boxes around three specific performance measures with distinct behaviors to guide the analysis (FPR , ACC , and P^2).

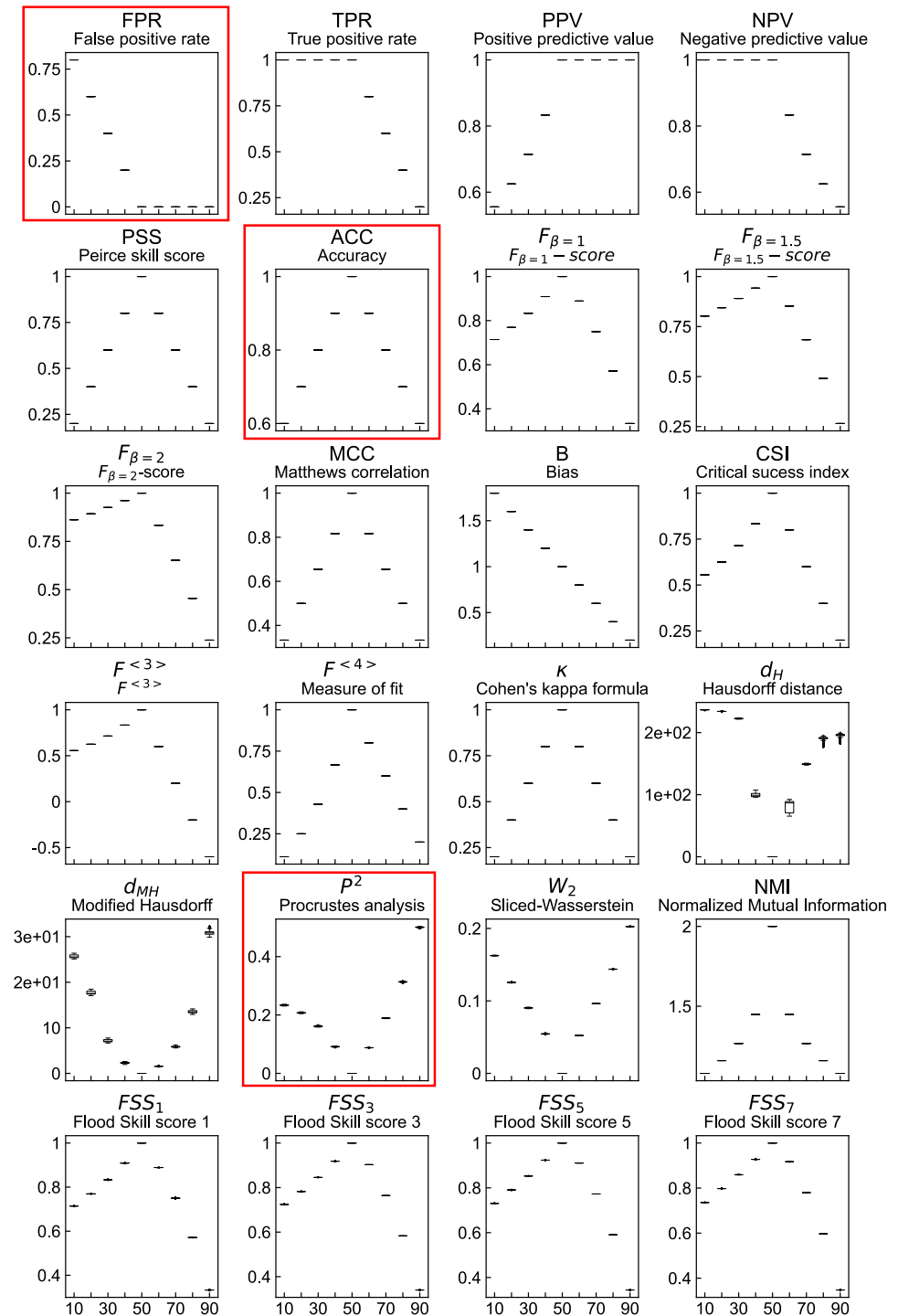


Figure 13. Box-plots of the evaluation of pixel-to-pixel and geometric performance measures for magnitude errors on 4,000 simulated flood maps. The reference flood maps extracted for a reference threshold (50th quantile) were compared to maps extracted with thresholds varying from the 10th to the 90th quantile of simulated water depths. Red boxes highlight performance measures with distinct behavior.

To pass the test, the quantity $|\mathcal{L}(S^{h_{50}}(\omega_m), S^{h_l}(\omega_m)) - \mathcal{L}(S^{h_{50}}(\omega_m), S^{h_{50}}(\omega_m))|$ was expected to increase when $|l - 50|$ increased. The optimal value of each performance measure was considered as $\mathcal{L}(S^{h_{50}}(\omega_m), S^{h_{50}}(\omega_m))$, as it is not always equal to zero. For instance, the *ACC* performance measure matches this, since $\mathcal{L}(S^{h_{50}}(\omega_m), S^{h_l}(\omega_m))$ becomes increasingly distant from its optimal value with increasing error. *FPR* did not pass the test, as it was insensitive for quantiles $h_l > h_{50}$ corresponding to smaller floods compared to the reference map. Another behavior emerged for some performance measures, such as P^2 , for which the numerical evaluation was not symmetric between negative and positive errors and tended to favor larger floods with higher values.

On our criterion, some performance measures failed the test. *FPR*, *TPR*, *PPV*, and *NPV* were not sensitive to increasing errors in flood extent for larger or smaller floods compared to the reference optimum. These results align with the definitions of these performance measures, which aim to minimize errors for either flooded or dry areas but not for both simultaneously. The other performance measures were sensitive to both smaller and larger floods, and can be considered as meeting the criterion. However, like P^2 , other performance measures (F_β , *CSI*, $F^{<3>}$, and *FSS*) were unbalanced, being more sensitive to the magnitude test for smaller or larger floods.

4.2.2. Observed Flood Maps

Application of this criterion to the observation described in Section 3.1 is shown visually in the Supporting Information S1. The behaviors associated with threshold errors were consistent with those observed for the criterion for the simulated fields. Similarly to the test on the simulated fields, threshold errors corresponded to magnitude errors. Flood extent was smaller in the observed flood maps than in the reference map when a negative threshold was applied and larger with a positive threshold.

4.3. Criterion 2: Sensitivity to Displacement Error

4.3.1. Translation Error

The results of the translation error experiment in the study case are shown in Figure 14. Special attention was paid to *FPR*, *ACC*, P^2 , and d_H , shown in red boxes, to focus on the distinct behavior of the performance measures of the experiment. The box-plots are ordered by increasing distance from the reference simulated fields. Performance measures passed the test if, on visual inspection, the trend was for monotonous decrease with respect to the reference values. While *FPR* and *ACC* deviated linearly from the optimal values with distance, d_H and P^2 were insensitive to certain translations. d_H had a constant value for any translation, and P^2 did not deteriorate for westward translation. Most performance measures (such as *FPR* and *ACC*) respected the criterion with better scores for shorter distances, exhibiting 3 distinct steps depending on whether they were translated by 1, 2, or 3 pixels. W_2 , d_H , and P^2 also failed the criterion, because of their non-monotonous behavior with increasing/decreasing distance.

4.3.2. Rotation Error

The results of the rotation error experiment are shown in Figure 15, with B and W_2 shown by red boxes representing all the behaviors of the test. The Sliced-Wasserstein distance W_2 met the criterion, since the value of $\mathcal{L}(S(\omega_m), S^\theta(\omega_m))$ deviated from its optimal value of zero with increasing rotation angle. All the performance measures had similar behaviors of increasing rotation error, except for bias B . $\mathcal{L}(S(\omega_m), S^\theta(\omega_m))$ was not monotonous with increasing rotation error, oscillating around its optimal value of 1 for rotation angles of 1° to 5° . d_H , P_2 , and W_2 were rejected by the criterion for translation error and B by the criterion for rotation error.

4.4. Criterion 3: Sensitivity to Noise

Results of the noise error experiment were analyzed for a kernel size of 9 with $s \in \{0.1, 0.2, \dots, 2.9, 3.0\}$ (Figure 16). Special attention was paid to *FPR*, B , d_H , and P^2 , highlighted with red boxes in the Figure, as they represent all the observed behaviors. P^2 showed a strictly monotonic trend with respect to the standard deviation. This is the expected behavior, since greater standard deviations mean that the observations are more noisy (more smoothing on pixel values) with respect to the reference. *FPR* was insensitive when s is lower than 0.5, failing the noise criterion as it could not detect the optimal maps (with the lowest noise) due to identifiability issues in these regions. The same issue arose for d_H , with jump discontinuities and identifiability problems at the various steps.

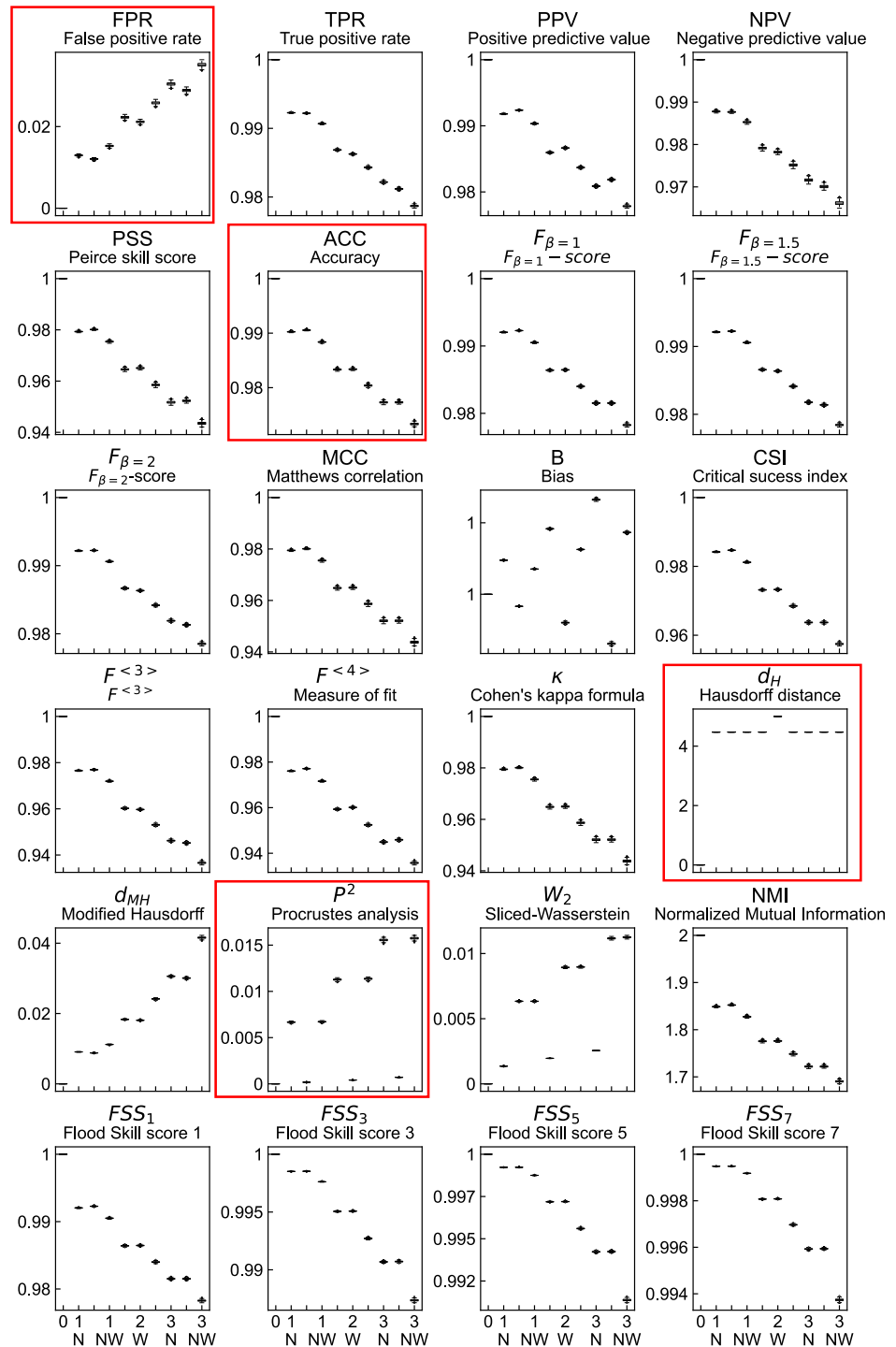


Figure 14. Box-plots of the evaluation of pixel-to-pixel and geometric performance measures for translation errors on 4,000 simulated flood maps. The reference flood maps were compared to translated maps (1–3 pixels in the north (N), northwest (NW) and west (W) directions). Red boxes highlight performance measures with distinct behavior.

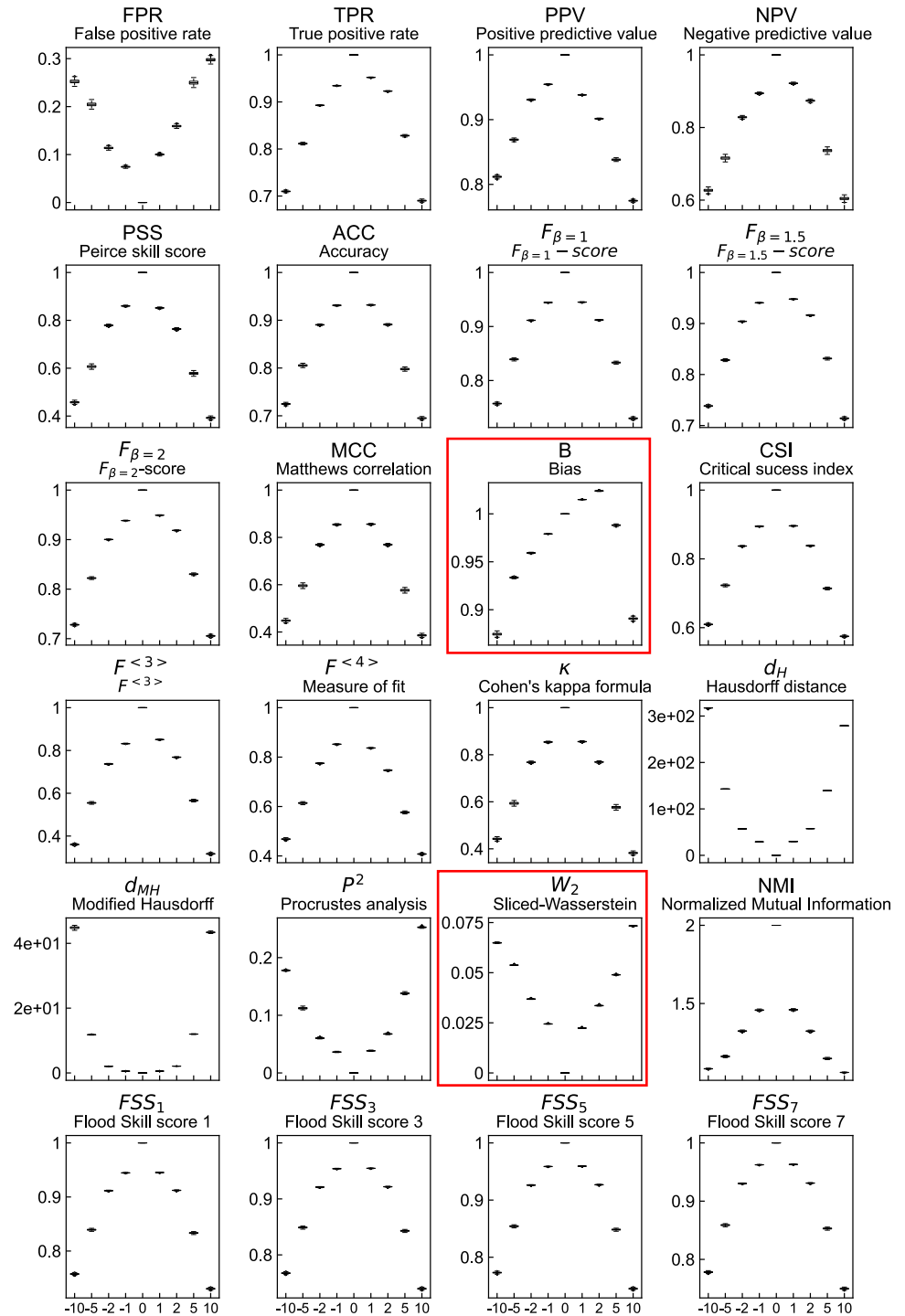


Figure 15. Box-plots of the evaluation of pixel-to-pixel and geometric performance measures for rotation errors on 4,000 simulated flood maps. The reference flood maps were compared to rotated maps (-10° – 10°). Red boxes highlight performance measures with distinct behavior.

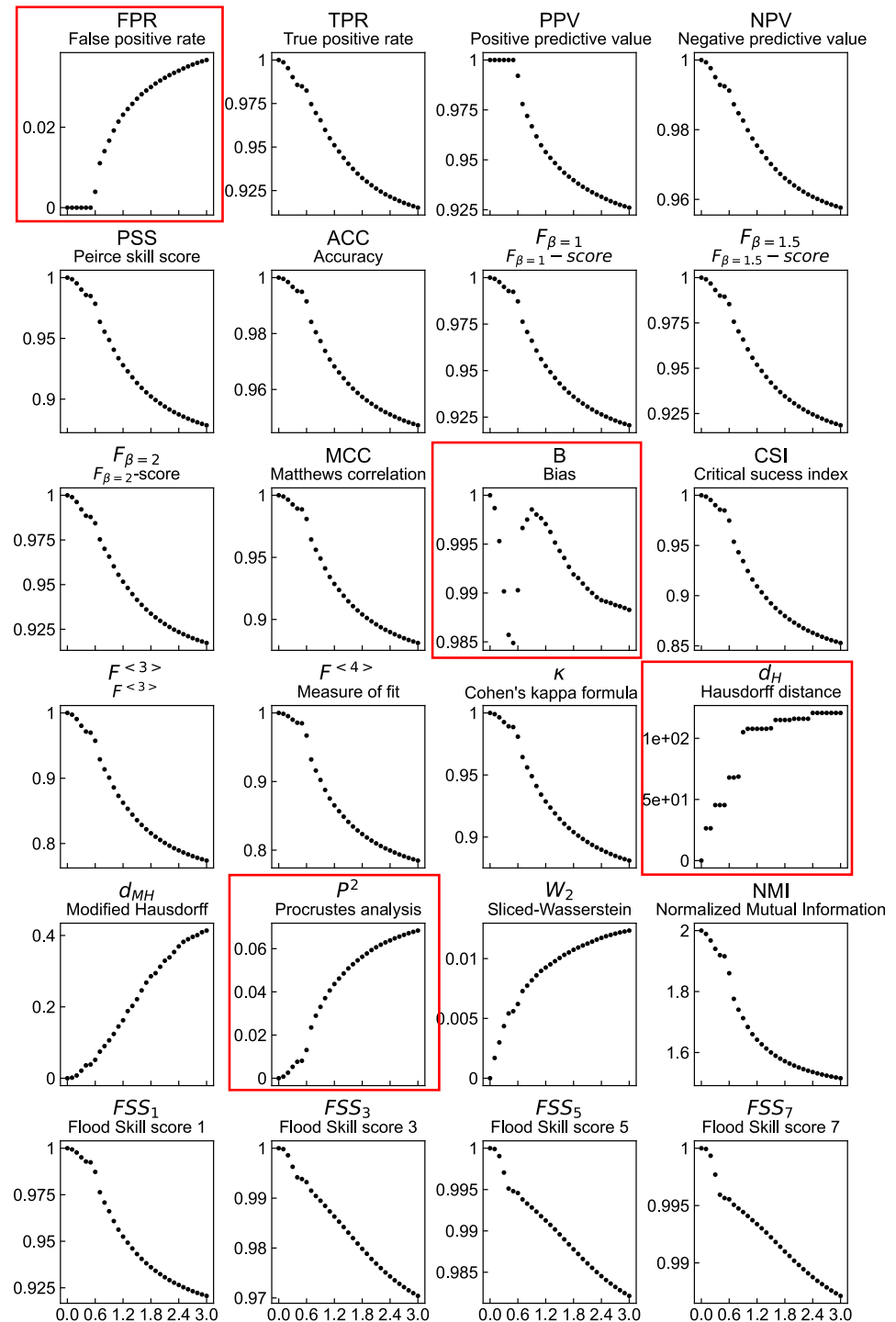


Figure 16. Evaluation of pixel-to-pixel and geometric performance measures for noise errors on the observed flood map. The reference flood map was compared to maps noised with a Gaussian filter with standard deviation varying from 0 to 3. Red boxes highlight performance measures with distinct behavior.

Table 4
Summary of the Metaverification Process for Pixel-To-Pixel and Geometric Comparisons

Pixel-to-pixel	FPR	TPR	PPV	NPV	PSS	ACC	$\beta = 1$	$\beta = 1.5$	$\beta = 2$	MCC	B	CSI	$F^{<3>}$	$F^{<4>}$	κ			
Magnitude	×	×	×	×	✓	✓	~	~	~	✓	✓	~	~	✓	✓			
Translation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Rotation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓			
Noise	×	✓	×	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓			
Time	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Geometric	d_H		d_{MH}		p^2		W_2		NMI		FSS_1		FSS_3		FSS_5		FSS_7	
Magnitude	✓		✓		~		✓		✓		~		~		~		~	
Translation	×		✓		×		×		✓		✓		✓		✓		✓	
Rotation	✓		✓		✓		✓		✓		✓		✓		✓		✓	
Noise	×		✓		✓		✓		✓		✓		✓		✓		✓	
Time	✓		✓		~		✓		✓		×		×		×		×	

Note. × = rejected by the criterion, ✓ = accepted by the criterion, and ~ = in between.

Another phenomenon occurred for B , for which performance was not monotonous with increasing standard deviation, consequently, maps with different noise levels exhibited the same value, which is not advisable. For this criterion, FPR and PPV were rejected because they were not sensitive to low noise levels, B because it was not monotonous, and d_H due to jump discontinuities.

4.5. Criterion 4: Computation Time

Computation time was assessed for $\mathcal{L}$ between two flood maps O and S . The performance measures were calculated on Intel(R) Xeon(R) Platinum 8260 CPU at 2.40 GHz processors (Cascade Lake). The performance measures based on a confusion matrix were fast to compute as they are based on simple element-by-element comparison of matrices (0.09 s). Cost differed between the others: FSS performance measures were costly (835 s) to calibrate a model on an iterative optimization approach, while, for the others, cost was low (3.00 s for d_H , 2.88 s for d_{MH} , 2.38 s for W_2 , 16.32 s for P^2 , and 0.67 s for NMI), except perhaps P^2 for special applications. In this study, FSS did not meet the criterion, and P^2 only partly met it, but this may have been due to inefficient optimization of the implementation or to the choice of algorithm to compute the performance measures.

4.6. Summary of the Selection Phase

The performance measures were evaluated for each criterion and are summarized in Table 4. The number of performance measures was reduced from 24 initially to 7, 17 being rejected for at least one of the criteria. The performance measures meeting all criteria are PSS , ACC , MCC , $F^{<4>}$, κ , d_{MH} , and NMI . The number of adapted performance measures is reduced. However, for calibrating or validating simulations with observations, the ideal is to have the fewest possible number of performance measures even in a multi-objective algorithm. As $F^{<4>}$ and ACC are linearly correlated (see Section 4.1), the ACC performance measure was excluded from the following analysis.

4.7. Ranking Performance Measures

The ranking process was tested on the 6 remaining performance measures. The flood maps were generated for thresholds σ^0 varying between -20 dB and -26 dB. For all performance measures, the $n_{\max}$ values were calculated. A rank-order emerged, with differences in $n_{\max}$ values: 60 for PSS , 76 for $F^{<4>}$, 95 for MCC , 96 for κ , and 100 for both d_{MH} and NMI . κ , MCC , NMI , and d_{MH} were the most robust with regard to errors in flood extent, and seemed more adequate for uncertain observations. For instance, for NMI , the same simulated map was optimal for the 100 uncertain observed flood maps. For two of the uncertain flood maps, the optimal simulated map for this performance measure was compared with the observations (Figure 17). The NMI value was higher when comparing the best simulation with the observed flood map extracted with $\sigma_{ref}^0 = -20$ dB.



Figure 17. Comparison of the best simulation for Normalized Mutual Information (NMI) with observed flood maps extracted with two thresholds: $\sigma_{ref}^0 = -20$ dB corresponding to $NMI = 1.07$ (left), and $\sigma_{ref}^0 = -26$ dB corresponding to $NMI = 1.05$ (right).

For the 100 observed flood maps, the simulation most often identified as close to observation differed between NMI , MCC , κ , and d_{MH} . The corresponding simulated maps are compared in Figure 18, with visualization of the simulated map for NMI and the difference in pixel values for the four performance measures in two zones of the study area where the front of the flood propagated.

The differences in the Figure are subtle, but, depending on the performance measure, the front of the flood in these two zones propagates over longer or shorter distances. For instance, in the 2 subdomains, front propagation was greater for the chosen simulated flood map for κ against MCC . Then, selecting performance measures may enable better representation of the flood in certain areas, which implies taking account of the economic/environmental priorities of the decision makers - one of the factors for the goodness of a simulation according to Murphy (1993).

5. Example of Calibration, and Discussion

After applying the methodology to a real flood event using a given flood model and one satellite observation, we provide an example of simple “calibration” of the model with one of the selected performance measures, and then discuss the methodology.

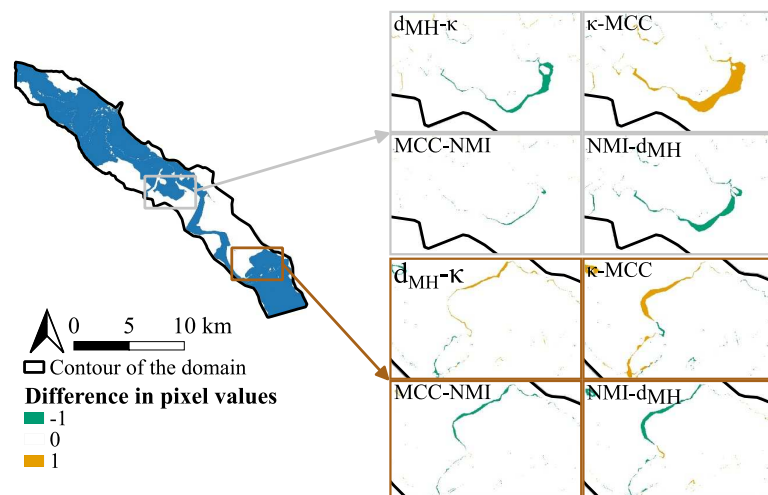


Figure 18. Optimized simulated flood map for Normalized Mutual Information (NMI) on the left, and comparison of the pixel values in two zones between four performance measures: Modified Hausdorff distance (d_{MH}), Cohen's kappa (κ), Matthew's correlation coefficient (MCC), and NMI .

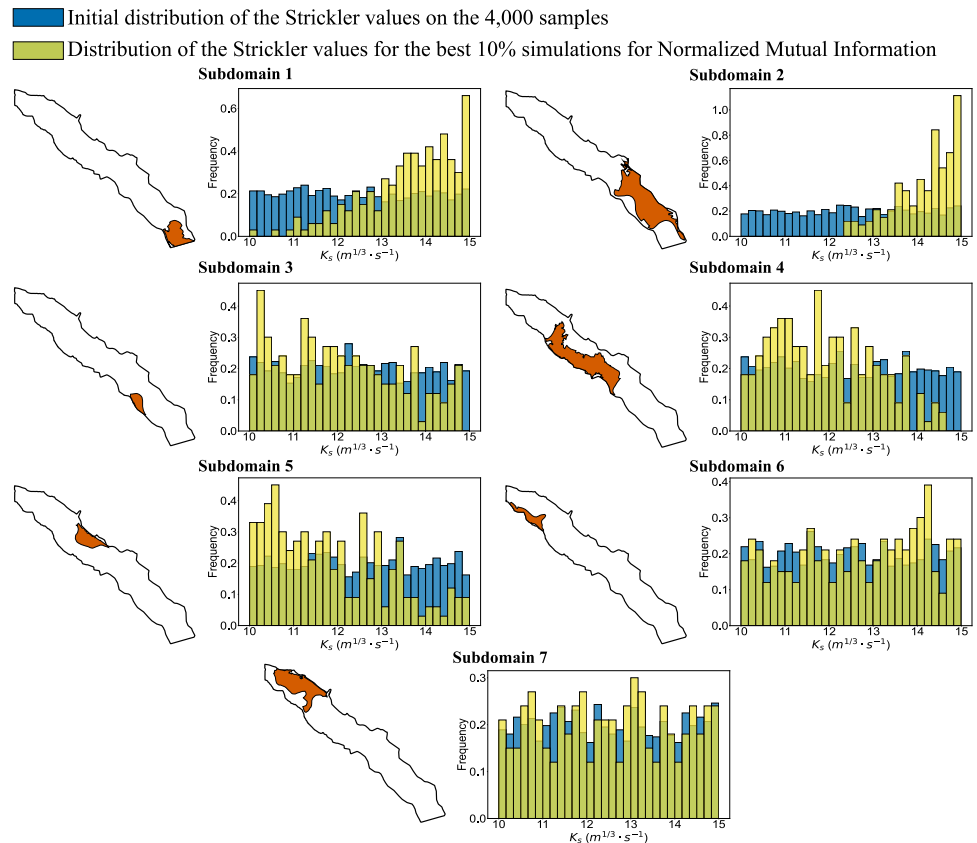


Figure 19. Calibration of 7 subdomains using Normalized Mutual Information by comparing the initial distribution of the friction K_s (on the 4,000 samples) with its posterior distribution for the best-performing simulations.

5.1. Calibration Example

As an illustration, the friction parameter in the seven largest subdomains was calibrated using one of the highly ranked performance measures, namely *NMI*. In Figure 19, the distribution of the friction parameter (Strickler values) for the 10% best simulations with respect to the *NMI* performance measure value is plotted against the initial reference distribution of the parameter for the 4,000 simulations. Comparison of simulations with observations on *NMI* indicated that the friction parameter was more likely in the upper zone of the prior distribution in subdomains 1 and 2. With the prior distribution of the Strickler values, the water depths in these subdomains are likely to be overestimated and the velocities underestimated (based on this observation). In contrast, in subdomains 4 and 5 the posterior distributions indicated that *NMI* was optimized for lower Strickler values. Using this performance measure, it is possible to infer the distribution of friction in these subdomains and improve the forecasting capacity of the model by automatic data assimilation, to calibrate the model based on observation.

While Figure 19 provides probability distributions of the parameter value in 7 subdomains for the best-performing simulations for *NMI*, it would also be useful to examine the optimal parameter values for the other selected performance measures and subdomains. Thus, Figure 20, shows the parameter fields for *MCC*, *NMI*, κ , and d_{MH} . The difference between the mean value of the parameter distribution for the best simulations and the mean value of the initial probability of the parameter field is shown for each performance measure. However, since the mean does not fully capture the characteristics of the entire probability distribution, we include complementary results in the Supporting Information S1 for the 10% and 90% quantiles of the distributions.

For the four performance measures, variations in the parameter field were consistent in almost all subdomains. For instance, optimal Strickler for the four performance measures was higher than the mean of the initial parameter

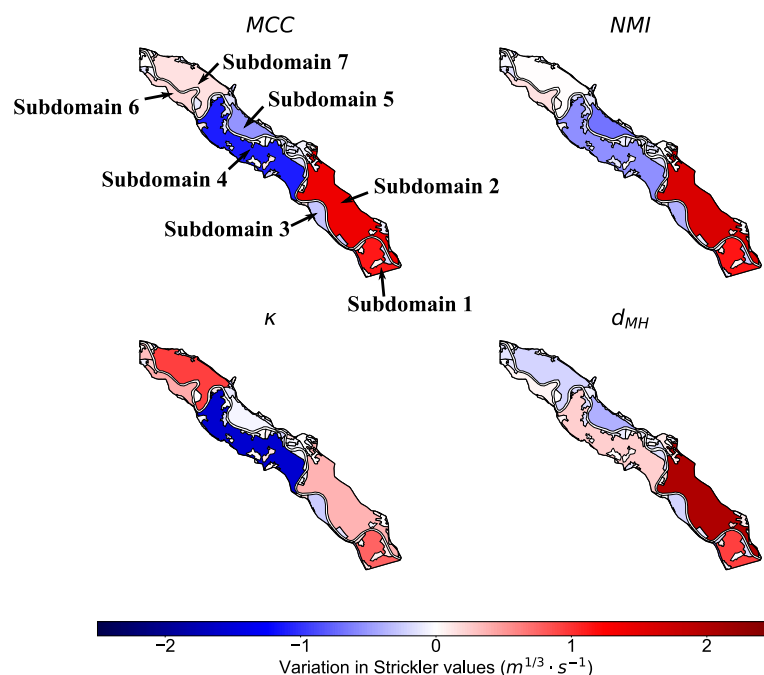


Figure 20. Difference between mean Strickler values within the 10% best simulations, and the mean of all Strickler values (in the 4,000 samples) for 4 performance measures.

value in subdomain 2, where, for all performance measures, the initial range of Strickler values was underestimated, leading to an overestimation of water depth and lower velocity, which could underestimate the simulated flood extent. However, d_{MH} behaved differently in subdomains 4 and 7: in subdomain 4, it was optimized for higher Strickler values, unlike the other performance measures, whereas conversely, in subdomain 7, d_{MH} was optimized for lower Strickler values compared to the other performance measures. Thus, if d_{MH} is chosen to calibrate this flood model, water depth will be lower in subdomain 4 and higher in subdomain 7 compared to the other performance measures. The parameter field for d_{MH} is more similar to that of NMI compared to the other performance measures. This similarity can be attributed to a stronger correlation between d_{MH} and NMI . The Figures in the appendices highlight the same phenomena at the 10th and 90th quantiles of the distributions.

5.2. Discussion and Limitations

The proposed methodology aims to provide a unified workflow for the practitioner to test performance measures using both pixel-to-pixel or geometrically based comparison. The criteria for the selection process were based on a literature review, hydraulic processes and satellite observation characteristics, but may not represent all factors for comparing a flood model against satellite observations. For instance, Landwehr et al. (2024) proposed several tests to quantify the impact of pixel sampling strategies on performance measures in imbalanced flood maps, where a disparity exists between flooded and dry pixels. Similarly, Pappenberger et al. (2007) and Stephens et al. (2014) examined the properties of pixel-to-pixel performance measures, focusing on their sensitivity to grid resolution and the influence of topography. These additional tests could be used to guide initial selection of the pixel-to-pixel performance measures, but are not adapted to geometric performance measures. However, the present study aimed to provide a comprehensive methodology to compare flood maps suited to any performance measure, whether pixel-to-pixel or geometric. Geometric performance measures could be useful for comparing flood maps, and should be assessed (Stephens et al., 2014); new ones have been recently reported such as the FSS for comparing flood maps (Hooker et al., 2022, 2023).

Pappenberger et al. (2007), Stephens et al. (2014), and Hooker et al. (2022) highlighted that the CSI performance measure (commonly used in the literature) leads to overestimation of flooded areas. This is consistent

with the results of the magnitude error experiment in the Garonne test case. Using the present criteria and ranking process, geometric performance measures such as Modified Hausdorff Distance and NMI were promising candidates for comparing observed and simulated flood maps in the test case. Future research should focus on evaluating geometric performance measures for comparing flood maps across a broader range of scenarios. Similarly, for pixel-to-pixel performance measures, MCC and κ were reliable for comparing the observed and simulated flood maps for the test case. Among the four performance measures, the pixel-to-pixel performance measures were faster to compute, and MCC or κ may be preferable in this test case. However, it could also be interesting to use more than one performance measure for multi-objective calibration, since we have seen that, for the best simulations, the front of the flood was slightly different depending on the performance measure.

The present study used observed flood maps extracted with a single extraction procedure (local thresholding). It could be investigated the impact of the extraction algorithm on the choice of performance measure, with the aim of determining whether any adjustments should be made to the methodology to consider other extraction algorithms. Local thresholding considers only unidirectional errors at a given time, whereas errors on flood delineation are more complex. Furthermore, the simulated fields were interpolated on the observation grid, leading to representativity errors between the simulated mesh and observed grid scales. An additional test may be needed for choosing of the performance measure, depending on these scales and the differences induced by mapping from a triangular mesh to a rectangular grid (Hooker et al., 2022). For the sensitivity to noise criterion, noise was considered uniform over the whole image, but could be adapted, since the uncertainty of flood mapping from SAR images is not always spatially uniform: for example, vegetated or urban areas are sometimes mistakenly considered flooded, while such errors are less likely in rural areas (Zhao et al., 2021). Instead of using a Gaussian filter that assumes spatially correlated noise, alternative noise models could be explored. Another means of adding relevant criteria would be to take account of the other two considerations referred to by Murphy (1993): practitioners' judgments of the simulations, and criteria on the economic/environmental priorities of the decision-makers.

Finally, the methodology is tailored to find a performance measure to compare an observation with a simulation, using a single scalar value for the whole domain. In optimization studies, practitioners may be more interested in obtaining local information to improve the spatial parameterization of their model. Choice of performance measure may be considered for local scales, rather than the whole image directly. For instance, the W_2 performance measure was rejected because of random projections on the grid, but may be more robust with other projections (e.g., along cross-sections of the river). Also, for comparison of the whole image, some performance measures were not tested; in the literature, hundreds of performance measures are available and could be tested such as performance measures based on fuzzy sets (Hagen, 2003) or feature-based approaches (Gilleland et al., 2009).

Additionally, it was noteworthy that the value range of the box-plots was narrow for most performance measures in Figures 13–15. This shows that a variation in flood extent generated with different magnitude or translation errors has greater impact than a change in the parameter field. The flood model used in this study was not very sensitive to the parameter field values, and it could be useful to apply the methodology to more sensitive test cases. However, as noted in Section 4.1, an optimal parameter field value exists for most performance measures. In any case, the present objective was to propose a methodology for practitioners to select a performance measure. Several performance measures were chosen, but practitioners are then free to test any other performance measure through this selection process.

6. Conclusion and Perspectives

There are many performance measures for comparing spatial fields with different specificities and properties, and the practitioner needs to be able to select the most reliable. The present study sets out a methodology to choose and rank performance measures to qualify a numerical flood model against satellite observations, especially in the case of extracted flood maps.

The methodology was applied to a study case for a flood event. The selection process highlighted that some performance measures used in the literature meet more criteria than others. In the present study case, most performance measures based on the confusion matrix were not equally sensitive to variations in flood extent, for both smaller and larger floods. The other criteria did not eliminate many performance measures, but each test still eliminated two or three. Analysis of the correlations showed that many performance measures were closely correlated, thus further reducing the number of performance measures when some are redundant. Finally, the

ranking process identified four performance measures (κ , MCC , NMI , and d_{MH}) respecting the present criteria and which were the least sensitive to observation errors in this test case. These four may not be the most suitable for other test cases; so the present study focused on proposing a methodology guiding performance measure selection in general.

To determine guidelines to generalize which performance measures should be used in a given context (e.g., depending on catchment or flood characteristics), the methodology should be tested on other flood events, hydrodynamic models, satellite observations; and flood map extraction procedures. The selection of a reliable performance measure in a data assimilation framework or to verify 2D simulated fields against 2D observations is essential, and we believe that taking time to analyze the choice of performance measure is important. In hydrology, with an increasing number of satellite missions, new data (e.g., water depth fields) become available for comparison against simulations. In the future, a similar study could be conducted to develop a methodology for selecting performance measures to compare observed and simulated water depth fields.

Appendix A: Extraction Procedure of the Flood Maps

Water and floods can be detected because open water surfaces have low backscattering values due to specular reflection and high water permittivity (Ulaby, 1982). One way to extract a flood map with pixels classified as dry or flooded is to determine a threshold that separates the two. A more straightforward approach is to utilize global thresholding, assuming that backscattering values for both dry and flooded pixels follow a Gaussian distribution. The best-known approach for determining the optimum threshold is the Otsu algorithm (Otsu, 1979), used, for instance by Schumann et al. (2009, 2010) and Pulvirenti et al. (2013). The algorithm maximizes the difference between the intensity values of the two classes while minimizing intra-class variance by iterating all possible threshold values. It relies on assumptions such as Gaussian distribution, and class sizes and variances should be both similar.

However, the bimodality of backscattering values is not always identifiable on the whole image, and the threshold cannot be determined directly. To address the spatial variability of the threshold for detecting water, Martinis et al. (2009), Twele et al. (2016), and Chini et al. (2017) proposed a split-based approach using a local-based thresholding algorithm. Local thresholding is based on splitting the image into sub-images with a bimodal histogram containing large portions of pixels belonging to flooded and dry classes (usually half water and half dry). Then, local thresholds are estimated and combined to calculate an overall threshold for the observation, leading to better results in flood extent detection compared to a global thresholding approach. Implementation in the present study was based on Chini et al. (2017). They used quad-tree decomposition on the satellite images and combined representative tiles with bimodal Gaussian distributions that contained at least 10% pixels of each class. The Ashman D coefficient, the Bhattacharyya coefficient and the surface ratio of the two classes were evaluated on each tile to evaluate the bimodality of the histogram, the Gaussian nature of the distributions, and the balance of each class. The Ashman D coefficient is formulated as:

$$AD(hist) = \sqrt{2} \frac{|\mu_1 - \mu_2|}{\sqrt{s_1^2 + s_2^2}}, \quad (A1)$$

where μ_1 , μ_2 , s_1 , and s_2 are the distribution means and standard deviations, respectively, and $hist$ is the histogram of the sub-tiles. The separability of the Gaussian distribution is ensured for high values of AD , and the present study and Chini et al. (2017) considered values above 2. The Bhattacharyya coefficient evaluating the similarity of two distributions is formulated as:

$$BC(hist, hist_f) = \sum_k \sqrt{hist(y_k)} \sqrt{hist_f(y_k)}, \quad (A2)$$

where y is the measurements, and k is the bins of the two discrete histograms $hist$ and $hist_f$ where $hist_f$ is the fitted histogram of $hist$. For a good fit of the normal distributions on the measured histograms, BC should be greater than 0.99. Finally, the third condition is that the smaller class should represent at least 10% of the second class.

Appendix B: Convergence of the Statistical Estimators

Let $Y = \{\mathcal{L}(O, S(\omega_m)) | 1 \leq m \leq M\} \in \mathbb{R}^M$ be a finite discrete random variable representing the values taken by $\mathcal{L}$ between an observation and M simulations. The statistical distributions of Y are analyzed through their moments. The first four moments (mean, standard deviation, skewness, and kurtosis) of the random variable are computed for different numbers of simulations. Moments of order k are defined as:

$$m_k = \sum_{m=1}^M \mathcal{L}(O, S(\omega_m))^k p(X = \mathcal{L}(O, S(\omega_m))). \quad (\text{B1})$$

With a sampling-based approach, estimators are approximated based on the available samples. To assess the robustness and convergence of the estimators, the bootstrap method is used to determine confidence intervals of the estimates. The number of bootstrap samples is set to 500 (after checking that the estimators confidence intervals do not change with more bootstrap samples). Then, the number of samples needed to reach stable statistical estimates is evaluated box-plot analysis on the median and confidence intervals of the estimators. An example of convergence analysis for ACC on the first fourth moment is available in the Supporting Information S1.

Appendix C: Correlation Analysis—Correlation Circle

Principal component analysis (PCA) is a dimensionality reduction technique that transforms the original variables, into a new set of orthogonal (uncorrelated) variables called principal components, which capture the maximum variance in the data. The variations of the performance measures for the first two principal components can be visualized on a correlation circle, where the lengths of the vectors represent the correlation between the variables and the principal components. Correlations between performance measures are more easily identified visually than in scatter plots, by looking at the angle between each vector representing the correlation between the variables. The correlations are visualized in the correlation circle in Figure C1 on the first two dimensions of the PCA (explaining 95% of the variance).

Contributions to dimensions differ depending on the performance measure. Except for d_H (red arrow in the Figure), all the variables are strongly correlated with the first two principal components. Linear correlations on the first two components can be identified by interpreting the angle between the arrows. To simplify visualization,

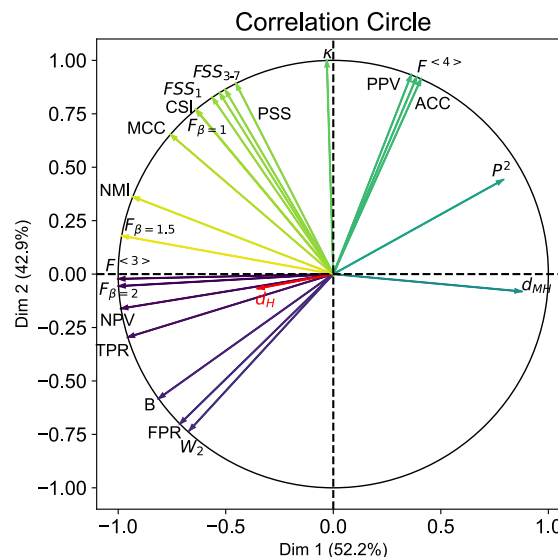


Figure C1. Correlation circle for the first two components of principal component analysis on the performance measures calculated for the reference flood map against 4,000 simulated flood maps.

closely correlated variables are shown with low variations in color: for example, NPV , $F_{\beta=2}$, and $F^{<3>}$ are closely correlated, as are PPV , $F^{<4>}$, and ACC .

Acronyms

2D	Two-Dimensional
ACC	Accuracy
B	Bias
CSI	Critical success index
FN	False Negative
FP	False Positive
FPR/F	False positive rate/Precision
FSS	Fraction Skill Score
F_{β}	F_{β} -score
$F^{<4>}$	Measure of fit
IGN	French National Institute of Geographic and Forest Information
MCC	Matthews correlation coefficient
d_H	Hausdorff distance
d_{MH}	Modified Hausdorff distance
κ	Cohen's Kappa formula
NMI	Normalized Mutual Information
NPV	Negative predictive value
PCA	Principal Component Analysis
PPV	Positive predictive value
P^2	Procrustes analysis
PSS	Peirce skill score
SAR	Synthetic Aperture Radar
SGS	Sequential Gaussian Simulations
TN	True Negative
TP	True Positive
TPR/H	True positive rate/Recall
W_2	Sliced-Wasserstein distance

Notation

O	Observed flood map
S	Simulated flood map
ω_m	Sample of the parameter
X, Y	Coordinates of the pixel centers



$\mathcal{L}$	Performance measure
N_x, N_y	Dimensions of the observed and simulated flood maps
N	Total number of pixels
M	Number of simulations
H	Simulated water depth field
σ^0	Backscattering intensity
σ_{ref}^0	Reference thresholding value of the backscattering
τ	Water depth thresholding value
s	Standard deviation
θ	Angle of rotation

Data Availability Statement

The openTELEMAC software suite is freely available on a Gitlab server with a track of all the developments, and a fixed branch for the v8p5 version used in the present study (https://gitlab.pam-retd.fr/otm/telemac-mascaret/-/tree/v8p5r0?ref_type=tags). For model construction, topographic data (for the Gironde and Lot-et-Garonne administrative départements) can be downloaded at <https://geoservices.ign.fr/rgealti>. For the land use data from 2018, the data set is available in a public catalog from the European Environment Agency (European Environment Agency, 2020). The hydrometric data at Tonneins, Marmande, and La Réole can be extracted from a public database (<https://www.hydro.eaufrance.fr/rechercher/entites-hydrometriques>) for stations O900001002, O909001001, and O919001001 respectively. The satellite observation used in the present study is available on ASF Data Search Vertex under scene name S1B_IW_GRDH_1SDV_20210202T175506_20210202T175531_025431_03077B_5B8F (<https://search.asf.alaska.edu/>). The Python libraries used for evaluating the performance measures (scikit-image, SciPy, Python Optimal Toolbox) are freely available (Flamary et al., 2021; van der Walt et al., 2014; Virtanen et al., 2020).

Acknowledgments

The present study was carried out as part of EDF R&D's MOISE-2 research project on river and coastal flood hazard assessment, whose support the authors gratefully acknowledge. It was supported and funded by the French National Association of Research and Technology (ANRT) and EDF R&D with the Industrial Conventions for Training through REsearch (CIFRE grant agreement 2022/0972). The authors would like to gratefully acknowledge the open-source community, especially that of the openTELEMAC and of the used Python libraries (scikit-image, SciPy, Python Optimal Toolbox). The authors also express special thanks to Sylvain Biancamaria for fruitful discussions on satellite imagery. Finally, we would like to thank the two anonymous reviewers whose comments and suggestions helped improve the manuscript.

References

- Aronica, G., Bates, P., & Horritt, M. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within glue. *Hydrological Processes*, 16(10), 2001–2016. <https://doi.org/10.1002/hyp.398>
- Aronica, G., Hankin, B., & Beven, K. (1998). Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources*, 22(4), 349–365. [https://doi.org/10.1016/S0309-1708\(98\)00017-7](https://doi.org/10.1016/S0309-1708(98)00017-7)
- Ballesteros, J., Bodoque, J., Díez-Herrero, A., Sanchez-Silva, M., & Stoffel, M. (2011). Calibration of floodplain roughness and estimation of flood discharge based on tree-ring evidence and hydraulic modelling. *Journal of Hydrology*, 403(1–2), 103–115. <https://doi.org/10.1016/j.jhydrol.2011.03.045>
- Barnes, H. H. (1967). *Roughness characteristics of natural channels (No. 1849)*. US Government Printing Office. <https://doi.org/10.3133/wsp1849>
- Besnard, A., & Goutal, N. (2011). Comparaison de modèles 1D à casiers et 2D pour la modélisation hydraulique d'une plaine d'inondation-cas de la garonne entre tonneins et la réole. *La Houille Blanche*, 97(3), 42–47. <https://doi.org/10.1051/lhb/2011031>
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1–2), 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>
- Bonneel, N., Rabin, J., Peyré, G., & Pfister, H. (2015). Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1), 22–45. <https://doi.org/10.1007/s10851-014-0506-3>
- Cao, Y., Ye, Y., Liang, L., Zhao, H., Jiang, Y., Wang, H., et al. (2019). A modified particle filter-based data assimilation method for a high-precision 2-D hydrodynamic model considering spatial-temporal variability of roughness: Simulation of dam-break flood inundation. *Water Resources Research*, 55(7), 6049–6068. <https://doi.org/10.1029/2018wr023568>
- Casati, B., Wilson, L., Stephenson, D., Nurmi, P., Ghelli, A., Pocerich, M., et al. (2008). Forecast verification: Current status and future directions. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, 15(1), 3–18. <https://doi.org/10.1002/met.52>
- Chini, M., Hostache, R., Giustarini, L., & Matgen, P. (2017). A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 6975–6988. <https://doi.org/10.1109/tgrs.2017.2737664>
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied hydrology*. McGraw-Hill Book Company.

- Cloke, H. L., & Pappenberger, F. (2008). Evaluating forecasts of extreme events for hydrological applications: An approach for screening unfamiliar performance measures. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, 15(1), 181–197. <https://doi.org/10.1002/met.58>
- Cooper, E. S., Dance, S. L., García-Pintado, J., Nichols, N. K., & Smith, P. J. (2019). Observation operators for assimilation of satellite observations in fluvial inundation forecasting. *Hydrology and Earth System Sciences*, 23(6), 2541–2559. <https://doi.org/10.5194/hess-23-2541-2019>
- Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S. P., Grimaldi, S., et al. (2015). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal*, 60(3), 402–423. <https://doi.org/10.1080/02626667.2014.903331>
- Dasgupta, A., Hostache, R., Ramsankaran, R., Schumann, G. J.-P., Grimaldi, S., Pauwels, V. R., & Walker, J. P. (2021). A mutual information-based likelihood function for particle filter flood extent assimilation. *Water Resources Research*, 57(2), e2020WR027859. <https://doi.org/10.1029/2020wr027859>
- Di Baldassarre, G., Schumann, G., & Bates, P. D. (2009). A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. *Journal of Hydrology*, 367(3–4), 276–282. <https://doi.org/10.1016/j.jhydrol.2009.01.020>
- Domeneghetti, A., Tarpanelli, A., Brocca, L., Barbetta, S., Moramarco, T., Castellarin, A., & Brath, A. (2014). The use of remote sensing-derived water surface data for hydraulic model calibration. *Remote Sensing of Environment*, 149, 130–141. <https://doi.org/10.1016/j.rse.2014.04.007>
- Dubuisson, M.-P., & Jain, A. K. (1994). A modified Hausdorff distance for object matching. In *Proceedings of 12th International Conference on Pattern Recognition* (Vol. 1, pp. 566–568). <https://doi.org/10.1109/icpr.1994.576361>
- El Garroussi, S., de Lozzo, M., Ricci, S., Lucor, D., Goutal, N., Goeury, C., & Boyaval, S. (2019). Uncertainty quantification in a two-dimensional river hydraulic model. In *International Conference on Uncertainty Quantification in Computational Sciences and Engineering*. <https://doi.org/10.7712/120219.6339.18380>
- European Environment Agency. (2020). Corine land cover 2018 (vector), Europe, 6-yearly - version 2020_20u1, May 2020. <https://doi.org/10.2909/71c95a07-e296-44fc-b22b-415f42acfd0>
- Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., et al. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78), 1–8. Retrieved from <http://jmlr.org/papers/v22/20-451.html>
- Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., & Ebert, E. E. (2009). Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5), 1416–1430. <https://doi.org/10.1175/2009waf2222269.1>
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), 33–51. <https://doi.org/10.1007/BF02291478>
- Grimaldi, S., Li, Y., Pauwels, V., & Walker, J. P. (2016). Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges. *Surveys in Geophysics*, 37(5), 977–1034. <https://doi.org/10.1007/s10712-016-9378-y>
- Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3), 235–249. <https://doi.org/10.1080/13658810210157822>
- Hall, J. W., Manning, L. J., & Hankin, R. K. (2011). Bayesian calibration of a flood inundation model using spatial data. *Water Resources Research*, 47(5). <https://doi.org/10.1029/2009wr008541>
- Hausdorff, F. (1927). *Mengenlehre* (Vol. 7). Walter de Gruyter.
- Hervouet, J.-M. (2007). *Hydrodynamics of free surface flows: Modelling with the finite element method*. John Wiley & Sons. <https://doi.org/10.1002/9780470319628>
- Hooker, H., Dance, S. L., Mason, D. C., Bevington, J., & Shelton, K. (2022). Spatial scale evaluation of forecast flood inundation maps. *Journal of Hydrology*, 612, 128170. <https://doi.org/10.1016/j.jhydrol.2022.128170>
- Hooker, H., Dance, S. L., Mason, D. C., Bevington, J., & Shelton, K. (2023). Assessing the spatial spread-skill of ensemble flood maps with remote-sensing observations. *Natural Hazards and Earth System Sciences*, 23(8), 2769–2785. <https://doi.org/10.5194/nhess-23-2769-2023>
- Horritt, M., & Bates, P. (2002). Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology*, 268(1–4), 87–99. [https://doi.org/10.1016/S0022-1694\(02\)00121-X](https://doi.org/10.1016/S0022-1694(02)00121-X)
- Horritt, M. S. (2000). Calibration of a two-dimensional finite element flood flow model using satellite radar imagery. *Water Resources Research*, 36(11), 3279–3291. <https://doi.org/10.1029/2000wr000206>
- Hostache, R., Chini, M., Giustarini, L., Neal, J., Kavetski, D., Wood, M., et al. (2018). Near-real-time assimilation of sar-derived flood maps for improving flood forecasts. *Water Resources Research*, 54(8), 5516–5535. <https://doi.org/10.1029/2017wr022205>
- Hostache, R., Lai, X., Monnier, J., & Puech, C. (2010). Assimilation of spatially distributed water levels into a shallow-water flood model. Part ii: Use of a remote sensing image of Mosel river. *Journal of Hydrology*, 390(3–4), 257–268. <https://doi.org/10.1016/j.jhydrol.2010.07.003>
- Hostache, R., Matgen, P., Schumann, G., Puech, C., Hoffmann, L., & Pfister, L. (2009). Water level estimation and reduction of hydraulic model calibration uncertainties using satellite SAR images of floods. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2), 431–441. <https://doi.org/10.1109/tgrs.2008.2008718>
- Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast verification: A practitioner's guide in atmospheric science*. John Wiley & Sons. <https://doi.org/10.1002/9781119960003>
- Landuyt, L., Van Wesemael, A., Schumann, G. J.-P., Hostache, R., Verhoest, N. E., & Van Coillie, F. M. (2018). Flood mapping based on synthetic aperture radar: An assessment of established approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 722–739. <https://doi.org/10.1109/tgrs.2018.2860054>
- Landwehr, T., Dasgupta, A., & Waske, B. (2024). Towards robust validation strategies for EO flood maps. *Remote Sensing of Environment*, 315, 114439. <https://doi.org/10.1016/j.rse.2024.114439>
- Manning, R. (1891). *On the flow of water in open channels and pipes* (Vol. 20, pp. 161–207). Transactions of the Institution of Civil Engineers of Ireland.
- Martinis, S., Twele, A., & Voigt, S. (2009). Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution Terrasar-X data. *Natural Hazards and Earth System Sciences*, 9(2), 303–314. <https://doi.org/10.5194/nhess-9-303-2009>
- Morvan, H., Knight, D., Wright, N., Tang, X., & Crossley, A. (2008). The concept of roughness in fluvial hydraulics and its formulation in 1D, 2D and 3D numerical simulation models. *Journal of Hydraulic Research*, 46(2), 191–208. <https://doi.org/10.1080/00221686.2008.9521855>
- Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8(2), 281–293. [https://doi.org/10.1175/1520-0434\(1993\)008<0281:wiagfa>2.0.co;2](https://doi.org/10.1175/1520-0434(1993)008<0281:wiagfa>2.0.co;2)
- Murphy, A. H. (1996). The Finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, 11(1), 3–20. [https://doi.org/10.1175/1520-0434\(1996\)011<0003:tfase>2.0.co;2](https://doi.org/10.1175/1520-0434(1996)011<0003:tfase>2.0.co;2)
- Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, 48(11). <https://doi.org/10.1029/2012wr012514>

- Nguyen, T. H., Ricci, S., Piacentini, A., Fatras, C., Kettig, P., Blanchet, G., et al. (2022). Dual state-parameter assimilation of SAR-derived wet surface ratio for improving fluvial flood reanalysis. *Water Resources Research*, 58(11), e2022WR033155. <https://doi.org/10.1029/2022wr033155>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. <https://doi.org/10.1109/tsmc.1979.4310076>
- Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R., & Matgen, P. (2007). Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. *Hydrology and Earth System Sciences*, 11(2), 739–752. <https://doi.org/10.5194/hess-11-739-2007>
- Pulvirenti, L., Marzano, F. S., Pierdicca, N., Mori, S., & Chini, M. (2013). Discrimination of water surfaces, heavy rainfall, and wet snow using cosmo-skymed observations of severe weather events. *IEEE Transactions on Geoscience and Remote Sensing*, 52(2), 858–869. <https://doi.org/10.1109/tgrs.2013.2244606>
- Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136(1), 78–97. <https://doi.org/10.1175/2007mwr2123.1>
- Sansosti, E., Berardino, P., Manunta, M., Serafino, F., & Fornaro, G. (2006). Geometrical SAR image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10), 2861–2870. <https://doi.org/10.1109/tgrs.2006.875787>
- Schumann, G., Di Baldassarre, G., Alsdorf, D., & Bates, P. (2010). Near real-time flood wave approximation on large rivers from space: Application to the river Po, Italy. *Water Resources Research*, 46(5). <https://doi.org/10.1029/2008wr007672>
- Schumann, G., Di Baldassarre, G., & Bates, P. D. (2009). The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8), 2801–2807. <https://doi.org/10.1109/tgrs.2009.2017937>
- Schumann, G., Giustarini, L., Tarpanelli, A., Jarihani, B., & Martinis, S. (2023). Flood modeling and prediction using Earth observation data. *Surveys in Geophysics*, 44(5), 1553–1578. <https://doi.org/10.1007/s10712-022-09751-y>
- Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., & Pfister, L. (2007). High-resolution 3-D flood information from radar imagery for flood hazard management. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6), 1715–1725. <https://doi.org/10.1109/tgrs.2006.888103>
- Schumann, G., Pappenberger, F., & Matgen, P. (2008). Estimating uncertainty associated with water stages from a single SAR image. *Advances in Water Resources*, 31(8), 1038–1047. <https://doi.org/10.1016/j.advwatres.2008.04.008>
- Stephens, E., Schumann, G., & Bates, P. (2014). Problems with binary pattern measures for flood model evaluation. *Hydrological Processes*, 28(18), 4928–4937. <https://doi.org/10.1002/hyp.9979>
- Studholme, C., Hill, D. L., & Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1), 71–86. [https://doi.org/10.1016/s0031-3203\(98\)00091-0](https://doi.org/10.1016/s0031-3203(98)00091-0)
- Tarpanelli, A., Mondini, A. C., & Camici, S. (2022). Effectiveness of sentinel-1 and sentinel-2 for flood detection assessment in Europe. *Natural Hazards and Earth System Sciences*, 22(8), 2473–2489. <https://doi.org/10.5194/nhess-22-2473-2022>
- Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: A fully automated processing chain. *International Journal of Remote Sensing*, 37(13), 2990–3004. <https://doi.org/10.1080/01431161.2016.1192304>
- Ulaby, F. T. (1982). *Radar remote sensing and surface scattering and emission theory*. Artech House.
- Ulaby, F. T., Kouyate, F., Brisco, B., & Williams, T. L. (1986). Textural information in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, (2), 235–245. <https://doi.org/10.1109/TGRS.1986.289643>
- van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., et al. (2014). Scikit-image: Image processing in Python. *PeerJ*, 2, e453. <https://doi.org/10.7717/peerj.453>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Waller, J. A., García-Pintado, J., Mason, D. C., Dance, S. L., & Nichols, N. K. (2018). Assessment of observation quality for data assimilation in flood models. *Hydrology and Earth System Sciences*, 22(7), 3983–3992. <https://doi.org/10.5194/hess-22-3983-2018>
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic Press. <https://doi.org/10.1016/C2017-0-03921-6>
- Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., et al. (2016). Calibration of channel depth and friction parameters in the LISFLOOD-FP hydraulic model using medium-resolution SAR data and identifiability techniques. *Hydrology and Earth System Sciences*, 20(12), 4983–4997. <https://doi.org/10.5194/hess-20-4983-2016>
- Zhao, J., Pelich, R., Hostache, R., Matgen, P., Cao, S., Wagner, W., & Chini, M. (2021). Deriving exclusion maps from C-band SAR time-series in support of floodwater mapping. *Remote Sensing of Environment*, 265, 112668. <https://doi.org/10.1016/j.rse.2021.112668>

WATER RESOURCES RESEARCH

Supporting Information for "Evaluation of Performance Measures for Comparing Flood Models with Satellite Observations"

J-P. Traver^{1,2}, S. Boyaval^{2,3}, C. Goeury^{1,2}, V. Bacchi¹, and F. Zaoui¹

¹EDF R&D, Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, France

²Laboratoire d'Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France

³Inria, Paris, France

Contents of this file

1. Figures S1 to S6

Introduction

This supporting information contains additional Figures referenced in the article.

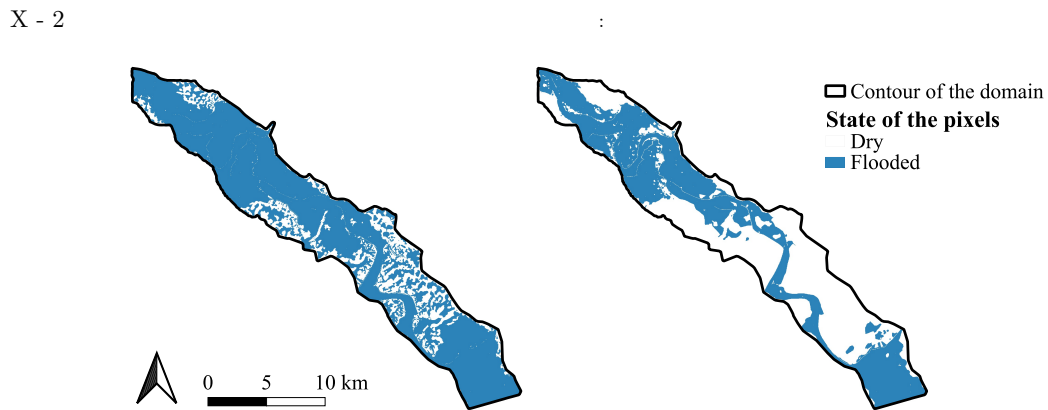


Figure S1. Example of a simulated flood map (on the Garonne River) for two thresholds: $h_{20} = 0.001 \text{ m}$ (left), and $h_{50} = 1.075 \text{ m}$ (right).



simulated flood maps, with r being the Pearson correlation coefficient

X - 4

:

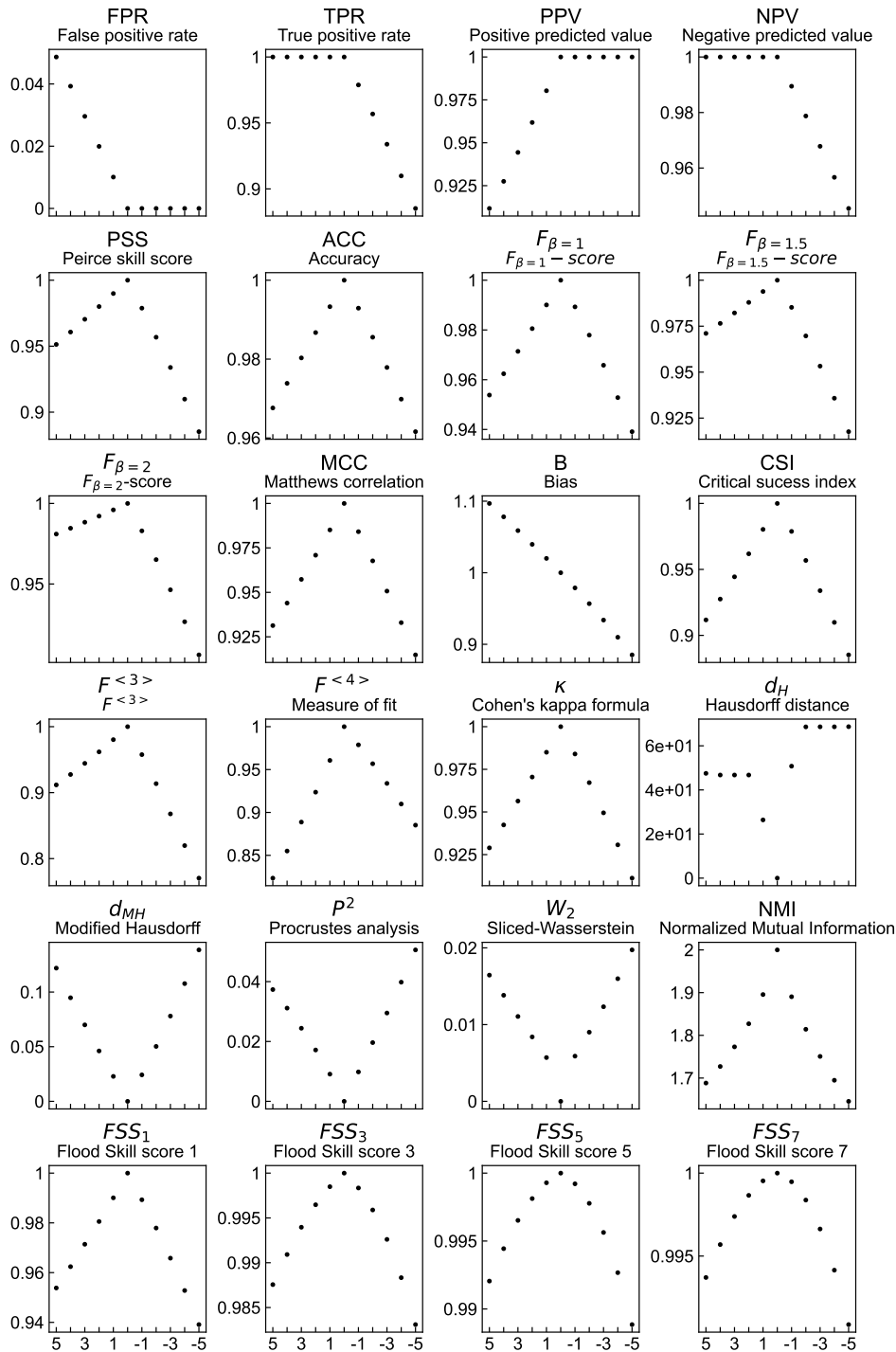


Figure S3. Evaluation of pixel-to-pixel and geometric performance measures for magnitude errors on the observed flood map. The reference flood map extracted for a reference threshold is compared to maps extracted with thresholds varying by -5% to +5% around the reference value.

X - 5

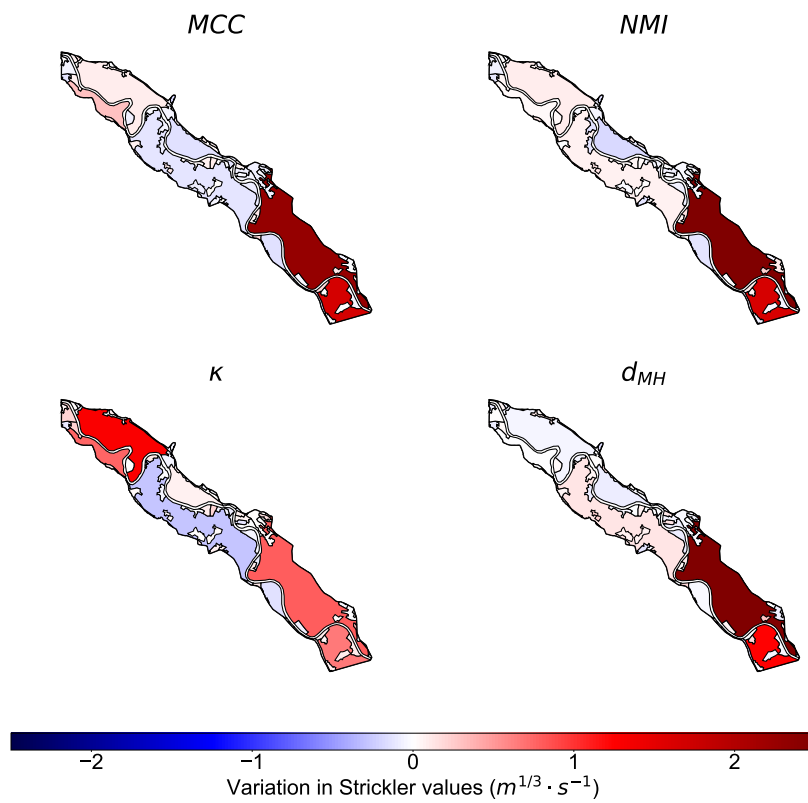


Figure S4. Difference between 10th quantile Strickler values within the 10% best simulations, and the 10th quantile of all Strickler values (in the 4,000 samples) for 4 performance measures.

X - 6

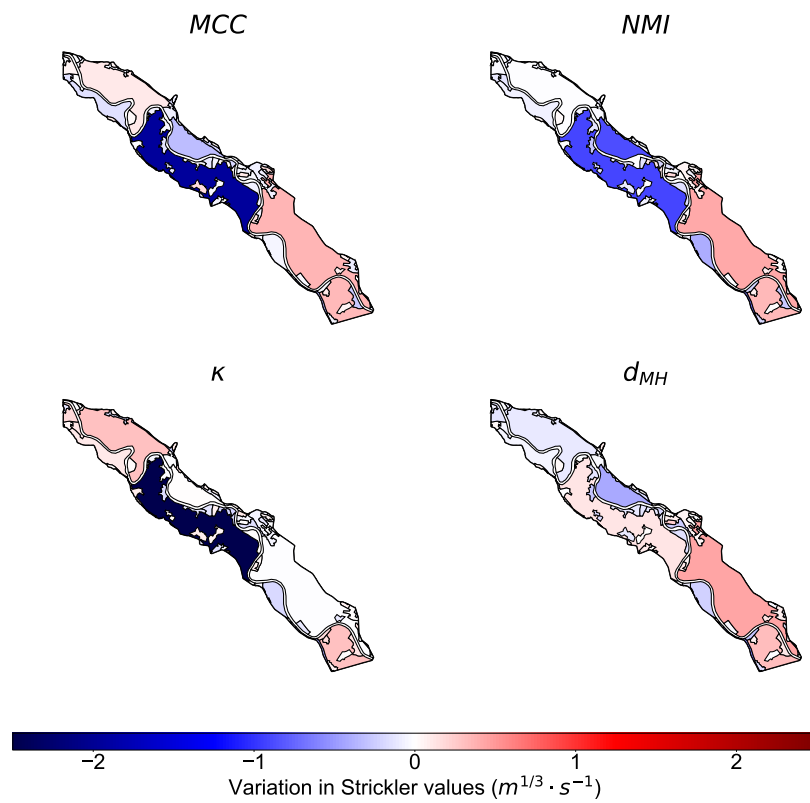


Figure S5. Difference between 90th quantile Strickler values within the 10% best simulations, and the 90th quantile of all Strickler values (in the 4,000 samples) for 4 performance measures.

X - 7

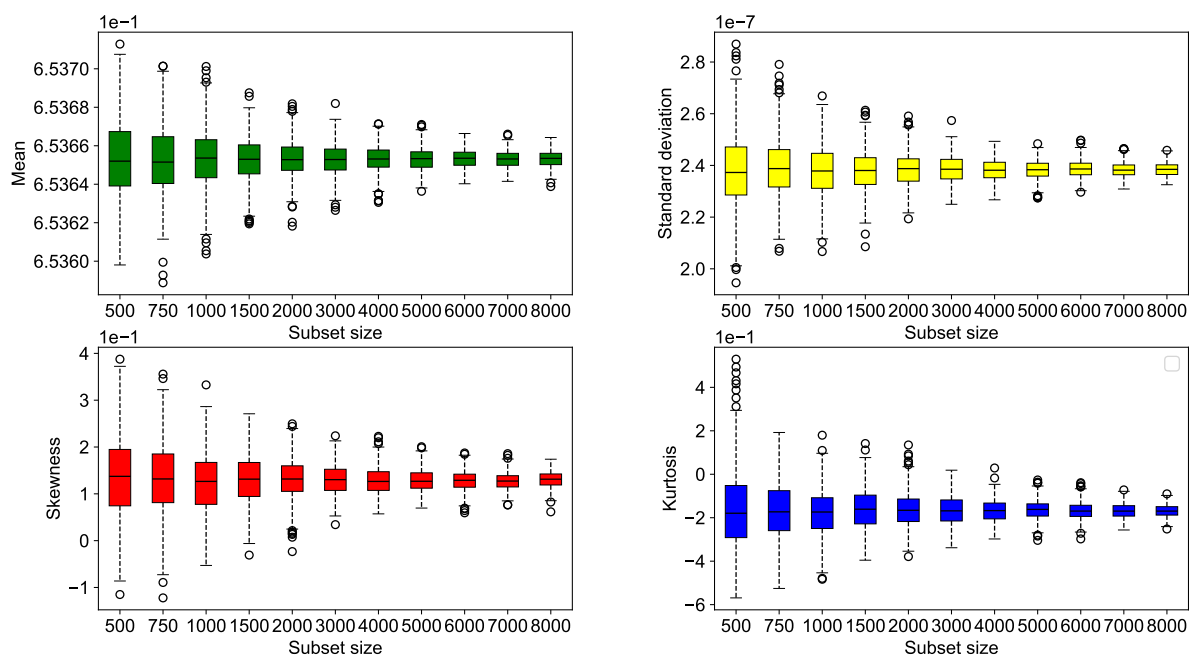


Figure S6. Mean, standard deviation, skewness and kurtosis evolutions for the Accuracy (ACC) performance measure on the bootstrap samples.

Chapter 6

Calibration of roughness field in 2D flood model with flood maps assimilation

Résumé :

Les modèles d'inondation reposent sur des équations basées sur des processus physiques, qui sont sujets à des incertitudes. L'une des principales sources d'incertitude est la distribution spatiale du paramètre de rugosité, qui doit être calibrée pour aligner les résultats simulés sur les données observées. Dans ce chapitre, le champ de rugosité est calibré à l'aide de cartes d'inondation issues de l'imagerie satellite, assimilées avec des techniques variationnelles. Dans ce cadre, une distance est nécessaire pour quantifier l'écart entre les cartes d'inondation observées et simulées. Ce chapitre étudie comment les différentes mesures de performance utilisées influencent le résultat de la calibration. L'analyse est effectuée sur un cas réel sur la Garonne en France; évalué à la fois avec une expérience jumelle et en utilisant des cartes d'inondation extraites d'images radars. Un cas complémentaire utilisant une rivière méandrée schématique avec trois sous-domaines de rugosité est étudié.

Abstract:

Flood models rely on physical process-based equations, which are subject to uncertainties. One key source of uncertainty is the spatial distribution of surface roughness, which must be calibrated to align simulated outputs with observed data. In this chapter, the roughness field is calibrated using flood maps from satellite imagery, assimilated with variational DA techniques. Within this framework, a distance is required to quantify the discrepancy between observed and simulated flood maps. This chapter investigates how different performance measures used in the variational DA framework influence the calibration outcome. The analysis is conducted on a real-world case on the Garonne River in France, evaluated both through a twin experiment and using flood maps extracted from Synthetic Aperture Radar observations. A complementary case using a schematic meandering channel with three roughness subdomains is studied.

Contents

6.1	Motivations	175
6.2	Preprint in preparation	175
6.3	Complements: Twin experiment on a schematic test case . .	203
6.4	Summary	203

6.1 Motivations

In order to reduce the uncertainty of hydrodynamic simulations, DA techniques can be used for parameter calibration. This chapter focuses on inferring spatially distributed roughness parameters by leveraging information extracted from satellite observations. The main contribution of this chapter is to address the third research question introduced in Chapter 1:

Research Question 3

Is it possible to calibrate floodplain roughness in 2D hydrodynamic models using uncertain satellite observations?

In Chapter 2, the importance of the roughness parameter in hydrodynamic modeling was discussed, along with a literature review on satellite imagery and mainly how SAR can be used to observe floods from space. In Chapter 4, we highlighted how to process SAR images to extract flood maps and evaluated the uncertainty in the processing pipeline from raw images to the flood maps. These flood maps are used as observations in this chapter to calibrate a 2D flood model. Moreover, we proposed in Chapter 5 a methodology to select a performance measure to compare simulated and observed flood maps. In this chapter, we evaluate the impact of these performance measures to assimilate the extracted flood maps as explained in Chapter 4 to infer a spatially distributed friction field.

Section 6.2 presents the assimilation of flood maps using different performance measures, which is evaluated on a real-world case, the Garonne River flood event, both through a twin experiment and using flood maps extracted from SAR observations. This section is written in the form of a paper that is in preparation for submission. Then Section 6.3 presents the assimilation of flood maps using different performance measures in a 2D flood model for a simple meandering channel configuration. This section presents the main results of this study on the schematic case. The associated paper can be found in the appendix and is written in the form of a conference paper presented at the Telemac User Conference.

6.2 Preprint in preparation

The following paper will be submitted to Computational Geosciences.

DATA ASSIMILATION OF FLOOD MAPS IN A 2D FLOOD MODEL FOR DIFFERENT PERFORMANCE MEASURES

Jean-Paul Travert

EDF R&D, Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, France
Laboratoire d'Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France
jean-paul.travert@edf.fr

Sébastien Boyaval

Laboratoire d'Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France
Inria, Paris, France

Cédric Goeury

EDF R&D, Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, France
Laboratoire d'Hydraulique Saint-Venant (LHSV), ENPC, Institut Polytechnique de Paris, EDF R&D, Chatou, France

Vito Bacchi

EDF R&D, Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, France

Fabrice Zaoui

EDF R&D, Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, France

ABSTRACT

This study explores the assimilation of uncertain flood maps derived from satellite observations into a 2D hydraulic model to calibrate a spatially varying roughness parameter, which is divided into subdomains. First, a sensitivity analysis is carried out to identify the most influential subdomains for various performance measure. Afterwards, Data Assimilation, specifically the three-dimensional variational method, is used on the reduced subset of parameters to calibrate the roughness field. Different cost function formulations are considered using various performance measures to compare the misfit between observed and simulated flood maps. Using synthetic and real observations, the study evaluates how these measures influence the calibration of model parameters and flood prediction accuracy. The application case is the Garonne River in France, where Synthetic Aperture Radar satellite images were available during flood events. The results indicate that the choice of performance measures in the cost function influences the calibration process. Furthermore, the study highlights that the current precision of derived flood maps is not sufficient for calibrating floodplain roughness.

1 Introduction

Flood simulations commonly rely on two-dimensional (2D) hydrodynamic models based on Shallow Water Equations (SWEs). These models must be carefully calibrated and validated using historical flood events to ensure accurate forecasting. Calibration involves adjusting model parameters to ensure that the simulated outputs match the observational data. Among these parameters, roughness coefficients, often expressed via the Strickler or Manning coefficients in operational flood studies, are crucial. These coefficients are frequently poorly known a priori, especially in floodplain areas. In operational contexts, model calibration is often performed through trial-and-error methods, using point-based measurements, such as time series from stage-gauging stations. However, these measurements are spatially sparse, with

missing information between the point-wise measurements. In the absence of observations in floodplains, roughness coefficients are typically prescribed by delineating homogeneous subdomains based on land use and assigning tabulated values from the literature (Barnes, 1967; Chow et al., 1988). This approach may oversimplify spatial variability and introduces uncertainty into flood predictions.

To overcome this limitation, spatially distributed data from remote sensing, such as 10 m resolution data from the Sentinel-1 satellite mission, provide new information for model calibration. The increasing availability of satellite missions now allows capturing "photographs" of the Earth during flood events. The satellite images can be leveraged to extract flood-related information that is difficult to obtain through in-situ measurements, including flood maps and water depth fields (Hostache et al., 2010). In particular, Synthetic Aperture Radar (SAR) imagery is now commonly exploited to derive flood maps, even under cloudy conditions (Grimaldi et al., 2016; Schumann et al., 2023). Flood maps can be combined with Digital Elevation Models (DEMs) to infer water depth fields. However, this process is prone to significant errors, on the order of meters, even when high-resolution DEMs are used (Oberstadler et al., 1997; Brakenridge et al., 1998; Cohen et al., 2018; Betterle and Salamon, 2024; Travert et al., 2025b). To reduce the uncertainties in water depth estimation, additional steps are often required. For instance, Hostache et al. (2009) introduced additional assumptions based on terrain characteristics and flow motions to constrain water depth estimation. Schumann et al. (2007, 2008) proposed a Regression and Elevation-based Flood Information eXtraction model (REFIX) to estimate water depth on cross-sections, considering a flat surface and regression techniques. However, assimilating flood maps is the most direct approach as it avoids introducing additional uncertainties deriving water depths (Cooper et al., 2019). Various mature techniques have been developed to extract flood maps from optical and SAR imagery (Chini et al., 2017; Otsu, 1979; Clement et al., 2018; Grimaldi et al., 2016) even if the processing of SAR images to generate flood maps is still subject to uncertainties (Schumann et al., 2008; Travert et al., 2025b).

Data assimilation offers a framework for optimally combining uncertain model outputs and observations (Carrassi et al., 2018). In particular, variational data assimilation has been successfully applied to estimate spatially distributed parameters in 2D flood models using remotely sensed data (Roux and Dartus, 2005; Lai and Monnier, 2009; Hostache et al., 2010; Lai et al., 2014; Larnier and Monnier, 2023). Variational data assimilation minimizes a cost function that measures the discrepancy between model outputs and observations, weighted by their respective uncertainties. One of the challenges in assimilating flood maps is selecting a suitable performance measure to quantify the discrepancy between the flood maps and model outputs (Lai et al., 2014; Dasgupta et al., 2021; Travert et al., 2025a). While for water depth fields assimilation the observed and simulated water depths are usually compared point to point weighted by their uncertainties (Matgen et al., 2010; Lai and Monnier, 2009; Hostache et al., 2010), flood maps, which are matrices of pixels with binary values (zeros and ones), can be more effectively compared to simulations using alternative performance measures.

In this study, a variational data assimilation framework was used to calibrate spatially varying roughness coefficients in a 2D flood model of the Garonne River in France, using flood maps derived from SAR observations. The impact of different performance measures on the calibration process is quantified, as well as the uncertainty introduced by observation errors. Travert et al. (2025a) proposed a methodology to select reliable performance measures for comparing simulated and observed flood maps. Based on this study on the same study area, we selected four performance measures, with two of them considered as suitable namely Cohen's Kappa formula and the modified Hausdorff distance, and two that were rejected by the methodology namely , the critical success index and Procrustes analysis. The methodology is validated through a twin experiment and then applied to real observations extracted during a flood event on the Garonne River in France in February 2021. Before data assimilation, a global sensitivity analysis is conducted to identify the influential parameters and reduce the dimensionality of the inverse problem.

The paper is organized as follows. Section 2 presents the study area, the hydrodynamic model, and the available observations. Section 3 describes the calibration framework, including sensitivity analysis and data assimilation using multiple performance measures. Section 4 reports the results of the sensitivity analysis, and Section 5 reports the results of the data assimilation step for both synthetic and real observations. Section 6 presents a discussion of the results and limitations of this study. Finally, Section 7 reports conclusions and perspectives.

2 Study area and material

This section presents the mathematical model governing the hydrodynamic evolution of two-dimensional free surface flows, along with the parameters that require calibration. Additionally, it introduces the observed flood maps used for calibrating the flood model. These flood map observations are derived from either a twin experiment or real satellite image observations.

2.1 Study area

In this work, the study area is a segment of the Garonne River in France of approximately 50 kilometers long, between the towns of Tonneins and La Réole (see Figure 1). The river reach width under typical flow conditions is approximately 250 meters, and the surrounding floodplain ranges in width from 1 to 4 km. The floodplains are mainly rural, consisting of agricultural lands and vegetated zones with limited urban development. Multiple dikes, installed following a significant flood event in 1875, protect urban settlements and infrastructure along the river. This segment of the Garonne River has been flooded multiple times and assessed in several flood modeling studies (Besnard and Goutal, 2011; Nguyen et al., 2022). This work builds upon prior investigations to calibrate and validate the numerical simulations.

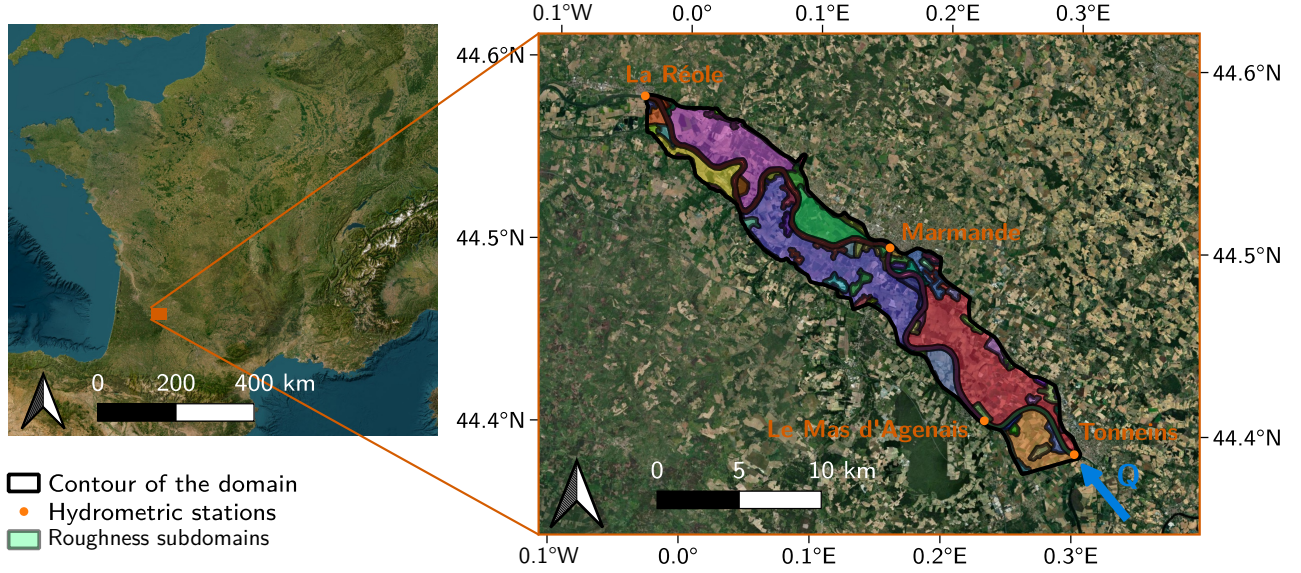


Figure 1: Study area on the Garonne River in France with the colored subdomains representing the roughness zones.

2.2 Satellite imagery

During February 2021, a flood event occurred on the Garonne River in France. The Sentinel-1 C-band Synthetic Aperture Radar (SAR) satellite mission acquired two observations of the study area. The SAR images were acquired on February 2, 2021, at 17:56 UTC, and on February 3, 2021, at 17:48 UTC. The “Ground Range Detected” (GRD) products were downloaded from the ASF Data Search Vertex portal (<https://search.asf.alaska.edu/>). In this study, the VH (vertical transmit, horizontal receive) polarization channel was used. The raw images consist of $N_x \times N_y$ pixel grids, with a spatial resolution of 10×10 m. Each pixel represents the backscattering coefficient (σ^0), which is proportional to the ratio of the radar energy received by the antenna to the energy transmitted. Flat water surfaces usually have low backscattering intensity due to specular reflection, making them easily distinguishable in SAR imagery.

2.3 Hydrodynamic model

The flood dynamics is modeled using the two-dimensional Shallow Water Equations (SWEs), presented in (1)–(3). These equations model the water flow over the domain Ω of Figure 1 equipped with Cartesian coordinates, assuming a uniform gravitational field $-g\vec{e}_z$, and are written as:

$$\frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} = 0, \quad (1)$$

$$\frac{\partial(hu)}{\partial t} + \frac{\partial(hu^2)}{\partial x} + \frac{\partial(huv)}{\partial y} = -gh \frac{\partial \eta}{\partial x} + \nabla \cdot (h \nu_e \nabla u) - \frac{1}{\rho} \tau_{bx}, \quad (2)$$

$$\frac{\partial(hv)}{\partial t} + \frac{\partial(hv^2)}{\partial y} + \frac{\partial(huv)}{\partial x} = -gh \frac{\partial \eta}{\partial y} + \nabla \cdot (h \nu_e \nabla v) - \frac{1}{\rho} \tau_{by}. \quad (3)$$

The system unknowns are the water depth $h = \eta - z_b \geq 0$, where η and z_b are the free-surface and bed elevations respectively, and the depth-averaged horizontal velocities $\mathbf{u} = [u, v]$. The latter functions are computed as solutions

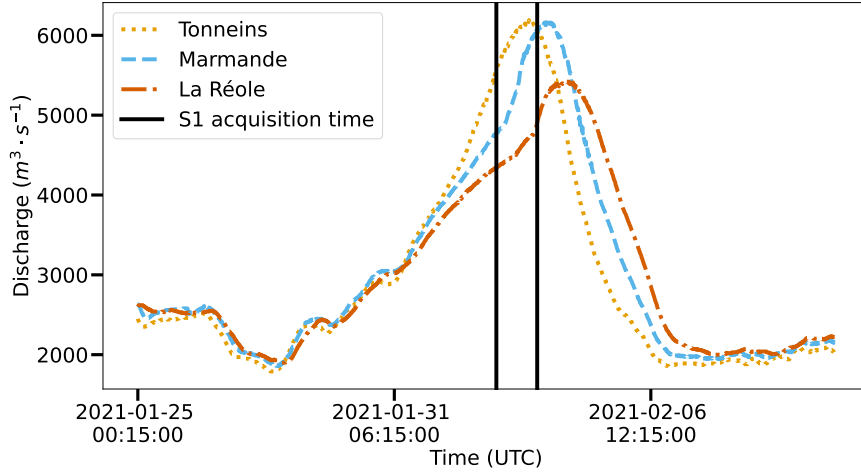


Figure 2: Measured discharge at the gauging stations (Tonneins, Marmande, and La Réole) and Sentinel-1 (S1) acquisition times during the February 2021 flood event.

to (1)–(3) when $t \in [0, T)$ and $(x, y) \in \Omega$ completed by conditions at $t = 0$ and wet boundaries $\partial\Omega \cap \{h > 0\}$. The effects of dispersion and diffusion are incorporated through a constant viscosity $\nu_e = \nu + \nu_t$, which includes the kinematic viscosity (ν), the eddy viscosity (ν_t). The bed shear stress vector τ , with components τ_{bx} and τ_{by} along the Cartesian coordinates (x, y) , represents the bed resistance to the flow. The bed shear stress is modeled as a function of h , u , and v to close the system, and is often written as:

$$\tau_b = \frac{1}{2} \rho C_f \mathbf{u} \sqrt{u^2 + v^2}, \quad (4)$$

where C_f is a non-dimensional friction coefficient determined with empirical or semi-empirical formulations (Morvan et al., 2008). In the present study, the Manning-Strickler formula (Manning et al., 1890) is used to model the friction coefficient, such that:

$$C_f = \frac{2g}{K_s^2 h^{1/3}}, \quad (5)$$

where K_s is the Strickler coefficient. The Strickler coefficient is assumed to be independent of time during flood events, but is considered dependent on (x, y) and requires calibration.

Equations (1)–(3) complemented with boundary conditions are solved with TELEMAC-2D (T2D) (Hervouet, 2007), which is part of the openTELEMAC numerical platform (www.opentelemac.org). The TELEMAC-2D solver uses a continuous, piecewise-linear finite element method based on an unstructured triangular mesh. Advection of velocity and water depth is computed using the N edge-by-edge and Positive Streamwise Invariant Distributive schemes, respectively. The wetting and drying of the mesh elements is handled through a corrected computation of the free surface slope (Hervouet, 2007).

In this study, the T2D model uses a triangular mesh composed of $\mathcal{N} = 41,416$ nodes. The mesh resolution ranges from 40 m in the river channel to 80 m near levees and up to 150 m in floodplain areas. Bathymetry within the channel was reconstructed using 70 cross-sectional elevation profiles, while the floodplain topography was derived from IGN (the French National Institute of Geographic and Forest Information) data and aerial imagery. A detailed description of the model setup is available in Besnard and Goutal (2011).

Initial conditions at $t = 0$ were set using a steady-state simulation corresponding to an upstream discharge of $2,300 \text{ m}^3/\text{s}$. The upstream boundary condition applied during the simulation was a time-dependent discharge curve measured at Tonneins (see Figure 2). Downstream, at La Réole, a rating curve was used to relate water depth to discharge. The flood event is simulated from the 25th of January 2021 to the 8th of February 2021.

The Strickler coefficient is piecewise constant across the subdomains. The riverbed was divided into three segments, each with a fixed Strickler coefficient based on previous model calibrations (Besnard and Goutal, 2011):

- $K_s = 45 \text{ m}^{1/3}/\text{s}$ between Tonneins and Mas d’Agenais,
- $K_s = 38 \text{ m}^{1/3}/\text{s}$ between Le Mas d’Agenais and Marmande,

- $K_s = 40 \text{ m}^{1/3}/\text{s}$ between Marmande and La Réole.

In floodplains, the Strickler coefficient varies in space depending on the land use. The domain is divided into 92 homogeneous areas, classified according to land use from the 2018 Corine Land Cover dataset (<https://land.copernicus.eu/en/products/corine-land-cover/clc2018>), as shown in Figure 1. Each land use class is associated with a specific range of plausible K_s values. These ranges are based on expert judgment and literature sources (Barnes, 1967; Chow et al., 1988). Table 1 reports the classical range of Strickler values used in 2D models in operational studies.

Table 1: Strickler coefficient ranges for different floodplain land use categories.

Land use	Strickler range ($\text{m}^{1/3} \cdot \text{s}^{-1}$)
Water bodies	$\mathcal{U}(32.5, 37.5)$
Fields and meadows without crops	$\mathcal{U}(17.5, 22.5)$
Croplands with low vegetation	$\mathcal{U}(15, 20)$
Croplands with tall vegetation	$\mathcal{U}(10, 15)$
Shrubland and undergrowth areas	$\mathcal{U}(8, 12)$
Low-density urban areas	$\mathcal{U}(8, 10)$
High-density urban areas	$\mathcal{U}(5, 8)$

2.4 Flood maps observations

In this study, two approaches were used to generate observed flood maps. First, synthetic observations were generated from a twin experiment for a controlled reference simulation. In a second approach, the observations were generated based on real SAR acquisitions described in Section 2.2. The flood maps observations are denoted as $O \in \{0; 1\}^{N_x \times N_y}$, i.e., matrices with entries in $\{0; 1\}$ (0 corresponds to dry pixels and 1 to flooded pixels). The two approaches for generating the flood maps are described in Sections 2.4.1 and 2.4.2.

2.4.1 Synthetic Observations from a Twin Experiment

Synthetic observations were generated using a reference Strickler configuration. For the channel, the reference values are described in Section 2.3. For the floodplains, the means of the intervals listed in Table 1 were used. Flood maps were then extracted from the reference simulation at two time points: February 2, 2021, at 17:56 UTC, and February 3, 2021, at 17:48 UTC, corresponding to the acquisition times of the real satellite observations. The simulated water depths at the two times of observation are projected on the reference grid of the satellite observations leading to $H \in \mathbb{R}^{N_x \times N_y}$ such that $H_{i,j}$ equals the value of the finite element water depth at the center of the pixel (i, j) . From this field, the flood map is created by applying a threshold τ to the water depth values such that $O_{i,j} = \mathbb{1}_{H_{i,j} > \tau}$. Typically, the threshold is fixed to 10 cm, following Aronica et al. (2002). Figure 3 reports the two extracted flood maps generated with the twin experiment.

2.4.2 Real Observations from Satellite Imagery

The second approach involves real satellite-derived flood maps. These maps are generated from Synthetic Aperture Radar (SAR) imagery (as described in Section 2.2), which provides measurements of the radar backscatter intensity, denoted σ^0 , for each pixel. The radar signal reflects differently off dry and flooded surfaces, allowing for classification. First, the speckle (noise) present in the SAR images is reduced using a deep-learning-based approach called SAR2SAR (Dalsasso et al., 2021). Then, using a local thresholding technique applied to the histogram of the SAR image, the preprocessed SAR image is segmented into a flood map (Martinis et al., 2015; Twele et al., 2016; Chini et al., 2017). According to Chini et al. (2017), the SAR images are decomposed into smaller sub-tiles to identify where the histogram of pixel intensity values is bimodal, the distributions are normally distributed, and both classes (flooded and dry pixels) are sufficiently represented. Then, a threshold σ_{ref}^0 is identified to separate dry from flooded pixels using an automatic segmentation approach, such as Otsu's threshold (Otsu, 1979). Pixels with $\sigma^0 < \sigma_{ref}^0$ are classified as flooded. This method is widely adopted in flood monitoring studies (Grimaldi et al., 2016; Landuyt et al., 2018; Schumann et al., 2023) and is also used operationally by institutions such as the Joint Research Centre for generating the Global Flood Monitoring product (Salamon et al., 2021). The flood maps extracted from the Sentinel-1 acquisition are reported in Figure 4.

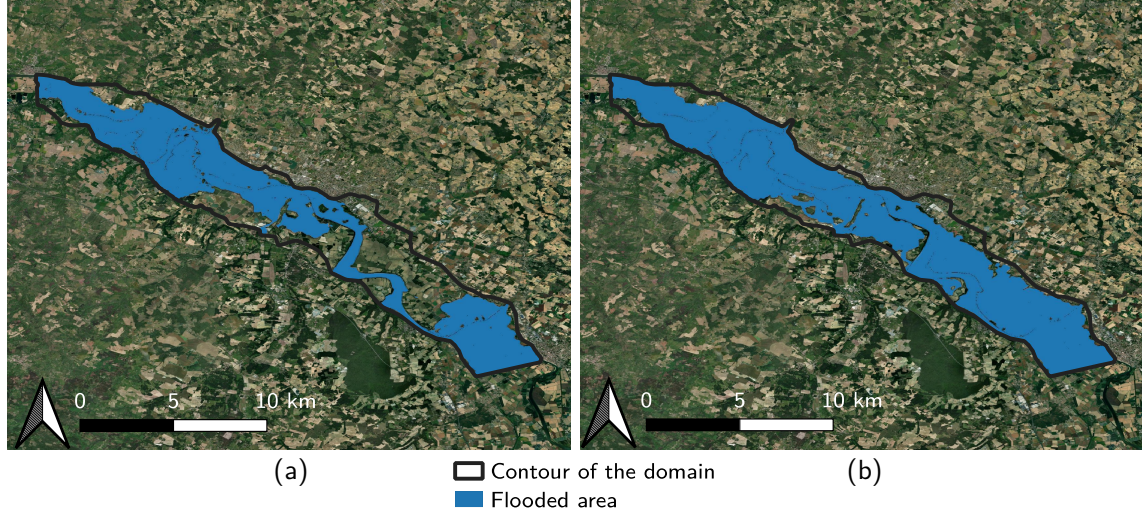


Figure 3: Synthetic flood maps observations generated with the twin experiment on the Garonne River using the reference roughness configuration on (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

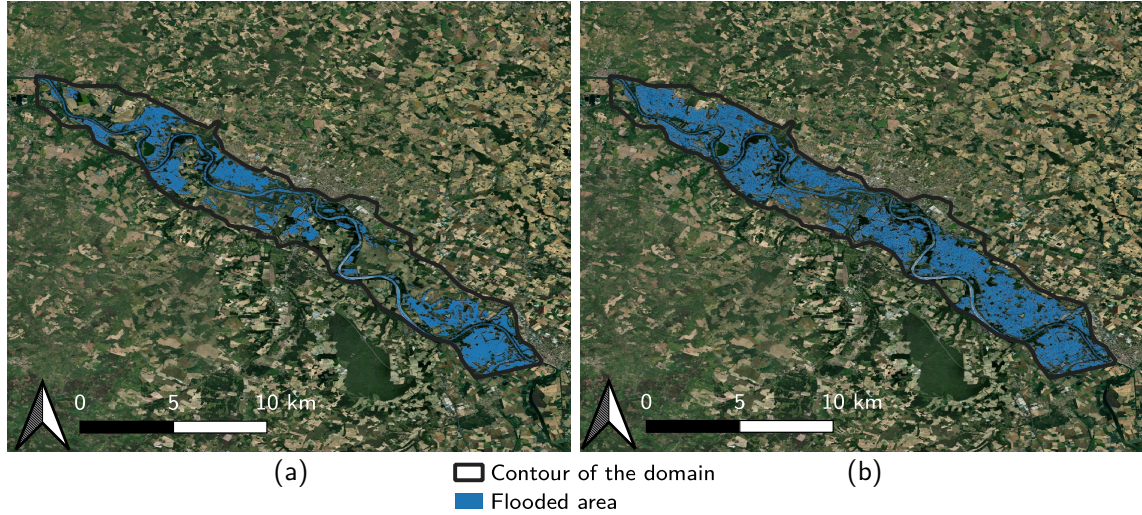


Figure 4: Flood maps extracted with the local thresholding method from Sentinel-1 acquisitions on the Garonne River on (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

3 Parameter estimation framework

Parameter estimation is a type of inverse problem, where unknown parameters are inferred from known outcomes such as observations. In this study, the goal is to estimate the spatially distributed roughness field θ based on observed flood maps $\mathbf{y}^{obs}$. Here θ is composed of the 92 floodplain subdomains described in Section 2.3, and $\mathbf{y}^{obs} = [O_1; O_2]$ is composed of two observed flood maps acquired at two distinct times. The parameter estimation problem is solved using a two-phase strategy, illustrated in Figure 5 and described below:

- Phase 1: A sensitivity analysis is carried out to select the most influential parameters among the 92 roughness parameters for each performance measure described in Section 3.2.3. The sensitivity analysis is based on Sobol's indices, which rank the parameters based on their contribution to the variance of the performance measure evaluation, comparing model outputs to observed flood maps.
- Phase 2: After selecting the influential parameters, the data assimilation step aims to infer the values of the selected parameters based on the available observations and background knowledge of the parameters. Using a variational approach, parameters are updated to minimize a cost function, which combines a background term

acting as a Tikhonov regularization (Tikhonov and Arsenin, 1977) and an observation term representing the misfit between observed and simulated outputs.

Each phase is described in detail in Sections 3.1 and 3.2.

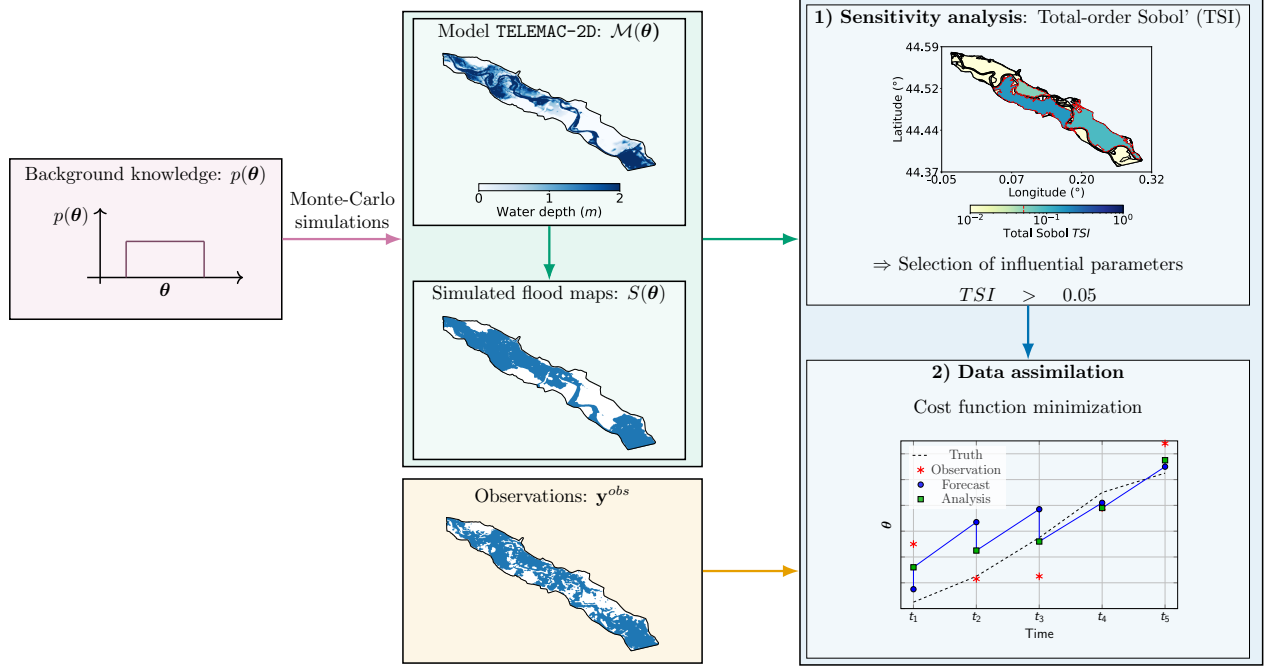


Figure 5: Methodology for calibrating roughness parameters in a two-dimensional flood model with observed flood map from satellite imagery.

3.1 Phase 1: Sensitivity analysis

The first phase aims to identify the most influential parameters among the 92 roughness parameters defined across the floodplain subdomains. This step reduces the computational cost and improves the robustness of the data assimilation phase by reducing the dimensionality of the parameter space. The sensitivity analysis phase is visualized in Figure 5 and described hereafter.

3.1.1 Algorithm

1. **Parameter sampling:** An ensemble of parameter vectors $\{\theta^{(k)}\}_{k=1}^M$ with M the number of samples is generated using the Latin Hypercube Sampling (LHS) strategy (McKay et al., 2000). The sampling bounds for each of the $N_p = 92$ parameters are listed in Table 1.
2. **Model execution:** For each sampled parameter vector $\theta^{(k)}$, the TELEMAC-2D model $\mathcal{M}$ is executed. The simulated water depths are projected onto the observation grid (grid of pixels of the observations) and transformed into flood maps at the times of observation using the observation operator $\mathcal{H}$. This operator maps the model state space (water depths on the simulated mesh) to the observation space (flood maps), as described in Section 2.4.1 considering flooded pixels for $h > 10$ cm. In our case, the application of $\mathcal{H} \circ \mathcal{M}$ results in two simulated flood maps denoted as $\mathbf{S} = [S_1; S_2]$ for the two times of observation.
3. **Performance measure evaluation:** The discrepancy between the simulated flood maps and one observation $O_j \in \mathbf{y}^{obs}$ (either from SAR imagery or a twin experiment) is quantified using a performance measure $\mathcal{L}$, such as:

$$Y_j^{(k)} = \mathcal{L}(O_j, S_j(\theta^{(k)})), \quad (6)$$

where $Y_j^{(k)}$ represents the performance measure evaluation associated with parameter vector $\theta^{(k)}$ for the j -th time of observation.

4. **Computation of Sobol' indices:** Total-order Sobol' indices $\{TSI_i\}_{i=1}^{N_p}$ are computed to quantify the contribution of each parameter to the variance of the output Y . The description of Sobol's indices computation is provided below.
5. **Selection of influential parameters:** Parameters with $TSI_i > 0.05$ are considered influential. Only these parameters are conserved for the data assimilation phase, leading to a reduced parameter space θ (restricted to the influential components).

3.1.2 Theory

Variance-based methods decompose the variance of output Y to quantify the relative importance of each input parameter θ_i (Saltelli et al., 2004). Among these methods, Sobol's indices (Sobol, 1993) are commonly used to quantify the contribution of the independent input $\{\theta_i\}_{1 \leq i \leq N_p}$ and their interactions in the variance $\mathbb{V}ar(Y)$. Based on the variance decomposition, the first-order and total-order indexes are written as follows:

$$SI_i = \frac{V_i}{\mathbb{V}ar(Y)} \quad (\text{First-order index}), \quad (7)$$

$$TSI_i = \frac{V_i + \sum_{j \neq i} V_{i,j} + \dots}{\mathbb{V}ar(Y)} \quad (\text{Total-order index}). \quad (8)$$

where $V_i(Y) = \mathbb{V}ar[\mathbb{E}(Y|\theta_i)]$, $V_{i,j}(Y) = \mathbb{V}ar[\mathbb{E}(Y|\theta_i, \theta_j)] - V_i(Y) - V_j(Y)$, etc. The term $\mathbb{E}(Y|\theta_i)$ represents the conditional expectation of the output Y under the assumption that the uncertain variable θ_i remains constant. SI_i which ranges between 0 and 1, measures the effect of θ_i alone, while TSI_i measures both direct and interaction effects with other parameters. Sobol's indices are typically computed using stochastic estimation of statistical moments via the Monte Carlo technique (Saltelli, 2002), but this approach becomes too computationally expensive in high dimensions. To address that computational-cost issue, Polynomial Chaos Expansion (PCE) was used, which approximates the model output Y . The PCE results in a surrogate model that enables a straightforward computation of Sobol's indices from the polynomial coefficients, as described in Sudret (2008). The confidence intervals for chaos Sobol' indices can be estimated by bootstrap as in Marelli and Sudret (2018). In the present study, Sobol's indices were computed using the OpenTURNS Python library (www.openturns.org) (Baudin et al., 2017). The confidence intervals of the estimation of the Sobol's indices and the distribution function of TSI_i supported that choosing $TSI > 0.05$ was adapted.

3.2 Phase 2: Data assimilation

The second phase aims to calibrate the influential roughness parameters identified during the sensitivity analysis (Phase 1 described in Section 3.1). This is achieved through a data assimilation approach as illustrated in Figure 5.

3.2.1 Algorithm

The main steps of the variational data assimilation phase are described in the following algorithm.

Algorithm 1 Parameter estimation via cost function minimization

- 1: Initialize parameters $\theta^{(0)} = \theta^b$ ▷ Background estimate
 - 2: $k \leftarrow 0$
 - 3: **while** $\Delta J \geq 10^{-2}$ and $k \leq 15$ **do**
 - 4: Run TELEMAC-2D with $\theta^{(k)}$ to obtain model output $\mathbf{S}(\theta^{(k)}) = [S_1; S_2]$
 - 5: For each observation, compute performance measure $Y_j^{(k)} = \mathcal{L}(O_j, S_j(\theta^{(k)}))$
 - 6: Compute cost function $J(\theta) = \frac{1}{2}(\theta - \theta^b)^T \mathbf{B}^{-1}(\theta - \theta^b) + \omega_1 \mathcal{L}(O_1, S_1(\theta))^2 + \omega_2 \mathcal{L}(O_2, S_2(\theta))^2$
 - 7: Update parameters $\theta^{(k)} \rightarrow \theta^{(k+1)}$ by minimizing J with gradient descent
 - 8: $k \leftarrow k + 1$
 - 9: **end while**
 - 10: **return** $\theta^{(k)}$ ▷ Estimated parameter vector
-

3.2.2 Theory

The variational data assimilation methods (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987) provides an optimal way of fitting a model to observations by combining uncertain observations and uncertain models. The

goal is to minimize a cost function that measures the mismatch between the model outputs and the observations, while incorporating background knowledge about the parameters. In this work, a variational approach, which is often used for parametric estimation problems was used (Hostache et al., 2010; Goeury et al., 2022). The cost function writes as follows:

$$\begin{cases} J(\boldsymbol{\theta}) = J_b(\boldsymbol{\theta}) + J_o(\boldsymbol{\theta}) \\ J_b(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^b)^T \mathbf{B}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^b) \\ J_o(\boldsymbol{\theta}) = \omega_1 \mathcal{L}(O_1, S_1(\boldsymbol{\theta}))^2 + \omega_2 \mathcal{L}(O_2, S_2(\boldsymbol{\theta}))^2 \end{cases}, \quad (9)$$

where $\boldsymbol{\theta}$ is the unknown parameter vector to calibrate, $\boldsymbol{\theta}^b$ is the background parameter estimate, $\mathbf{B}$ is the background error covariance matrix, and ω_1 and ω_2 are weights adjusting the relative influence of the observation terms in the cost function. In the present study, these weights are taken equal and simplified as ω . The first term regularizes the problem (similar to Tikhonov regularization), and the second quantifies the misfit between the simulated and observed flood maps. The performance measures used in this study are described in Section 3.2.3. Considering the dependence of flood maps as smooth in a manifold equipped with distance $\mathcal{L}$, J can be considered continuous and piecewise differentiable in $\boldsymbol{\theta}$.

The minimization of J was carried out using the Limited-memory Broyden Fletcher Goldfarb Shanno (L-BFGS) quasi-Newton method (Byrd et al., 1995). In this minimization, the gradient of J can be computed using an adjoint model or via finite differences. In the present study, the gradient of J with respect to the parameters was computed through finite differences, such as:

$$\nabla_j J(\boldsymbol{\theta}) \approx \frac{J(\boldsymbol{\theta} + \epsilon \theta_j \mathbf{e}_j) - J(\boldsymbol{\theta})}{\epsilon \theta_j}, \quad (10)$$

where $\mathbf{e}_j$ is the unit vector in the j -th direction and ϵ is the differentiation step. The differentiation step is set to 10^{-3} . The differentiation step should be sufficiently small to approximate the gradient.

The optimization proceeds iteratively, updating the parameter estimate until the convergence criteria are satisfied (e.g., a reduction in J or changes in $\boldsymbol{\theta}$ below a given threshold, or a maximum number of iterations). In this study, the maximum number of iterations of the algorithm was fixed to 15 to limit the computation time. When using finite differences to compute the gradient each iteration corresponds to $N_p + 1$ simulations. In this study, the simulation time for one simulation was approximately 30 minutes on 36 Intel Xeon Gold 6140 processors. To avoid unnecessary iterations on nearly identical states, a maximum relative cost reduction in J was set to 10^{-2} . The variational approach was used through ADAO (<https://pypi.org/project/adao/>), a Python data assimilation library (Argaud, 2019).

3.2.3 Performance measures

Travert et al. (2025a) proposed a methodology to guide the selection of performance measures for comparing numerical flood simulations with satellite imagery to improve model calibration and validation. Using the Garonne River as a case study, they evaluated 24 potential performance measures and identified four as the most suitable for calibrating the flood model. In the present study, two of the selected performance measures are used: the Cohen's Kappa formula and the modified Hausdorff distance. Additionally, two other performance measures that were not selected as reliable were used: Procrustes analysis and the critical success index (also known as the measure of fit). These four performance measures are described below. The critical success index and Cohen's Kappa formula are based on pixel-to-pixel comparisons, while the Procrustes analysis and modified Hausdorff distances compute geometric similarity via contour-based spatial distances.

Pixel-to-pixel comparison At each grid point, a hit corresponds to a correctly predicted flooded pixel known as True Positive (TP), while a correct rejection or True Negative (TN) occurs when a dry pixel is accurately identified as dry. A false alarm, known as a False Positive (FP), arises when the model predicts flooding at a dry location, and a miss, known as a False Negative (FN), occurs when a flooded area is incorrectly classified as dry in the simulation. In this study, the Cohen's Kappa formula and the critical success index are used:

- Cohen's Kappa formula measures the agreement between the simulated and observed binary classifications while accounting for agreement occurring by chance. It is defined as:

$$\kappa = \frac{2 \cdot (TP \cdot TN - FN \cdot FP)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)}. \quad (11)$$

A κ of 1 indicates perfect agreement, while 0 indicates no agreement beyond chance.

- The critical success index, also known as the measure of fit, quantifies the proportion of correctly predicted flooded pixels out of all pixels where flooding was predicted or observed and is defined as:

$$CSI = \frac{TP}{TP + FP + FN}. \quad (12)$$

The CSI ranges from 0 (no prediction skill) to 1 (perfect prediction).

Geometric comparison An alternative approach is to compare geometric properties of the simulated $S \in \{0, 1\}^{N_x \times N_y}$ and observed $O \in \{0, 1\}^{N_x \times N_y}$ flood maps. Two of these alternatives are presented with Procrustes analysis and the modified Hausdorff distance.

- Procrustes analysis (Gower, 1975) is a shape alignment technique that compares two shapes by eliminating differences in position, scale, and orientation, as visualized in Figure 6.

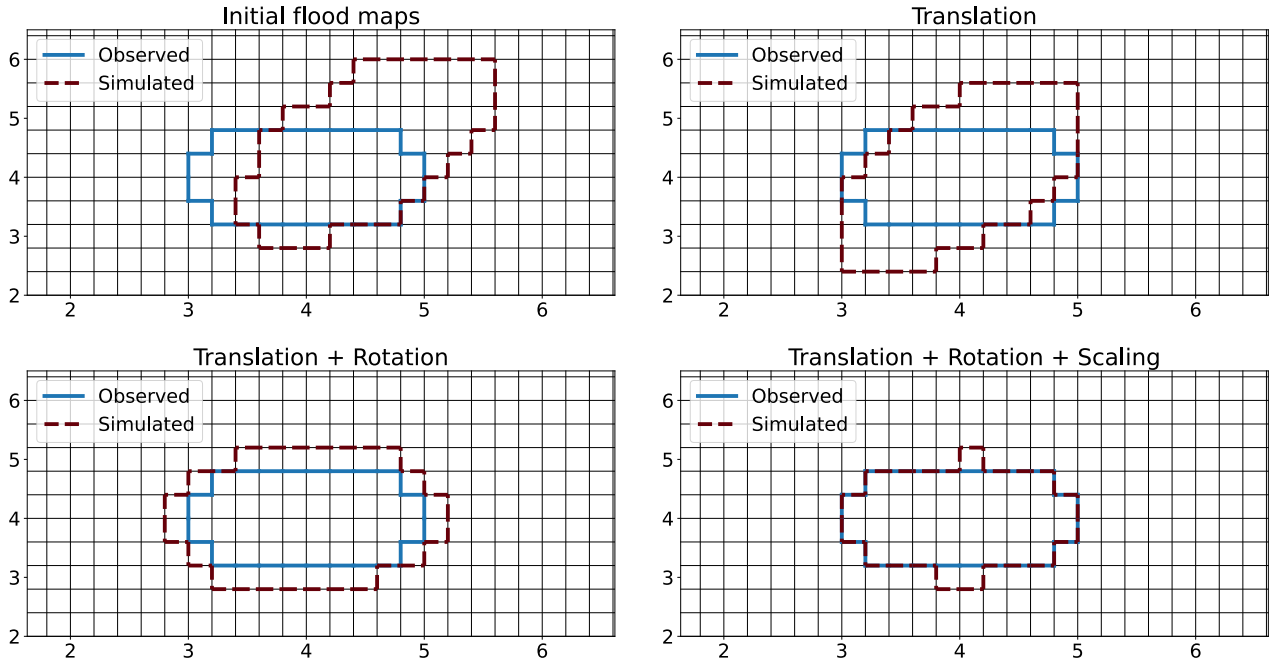


Figure 6: Optimal transformation between two schematic flood maps achieved by translating, rotating, and scaling the simulated flood map (brown contour) extracted from (Travert et al., 2025a).

The original O and S matrices are transformed into two matrices A_1 and A_2 by subtracting their mean, and such as $Tr(A_i \cdot (A_i)^T) = 1$ with $i \in \{1, 2\}$. Then, an optimal transformation is represented by an orthogonal matrix Q and a scaling factor c , which are determined using singular value decomposition (SVD). The aim is to minimize the Procrustes distance P^2 , computed as:

$$P^2 = d_e(A_1, cA_2Q), \quad (13)$$

where

$$d_e(A_1, cA_2Q) = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (A_1)_{i,j} - (cA_2Q)_{i,j})^2}. \quad (14)$$

- The modified Hausdorff distance (Dubuisson and Jain, 1994) addresses the sensitivity of the classical Hausdorff distance to outliers by replacing the maximum operator with an average, such as:

$$d_{MH}(O, S) = \max \left\{ \frac{1}{|O|} \sum_{(i,j) \in O} \min_{(k,l) \in S} \delta((i,j), (k,l)), \frac{1}{|S|} \sum_{(k,l) \in S} \min_{(i,j) \in O} \delta((i,j), (k,l)) \right\}, \quad (15)$$

where $\delta((i,j), (k,l))$ is the Euclidean distance between location of pixels (i,j) and (k,l) . $|O| = \sum_{i,j} |O_{i,j}|$ and $|S| = \sum_{i,j} |S_{i,j}|$ represent the number of flooded pixels in the observation and simulation, respectively. w

The parameter estimation framework described in this Section is applied on flood event described in Section 2. The results of the sensitivity analysis step of the methodology are described in Section 4, and then are exploited in the data assimilation step in Section 5.

4 Results of the sensitivity analysis step

First, the sensitivity analysis is carried out following the framework described in Section 3.1. The sensitivity analysis step was conducted on the twin and real experiment. Only the results from the twin experiment setup are reported here, as the influential subdomains were identical in the twin and real experiments.

On the twin experiment, a sensitivity analysis was carried out to identify the influential parameters on the scalar output $\mathcal{L}(O_j, S_j(\theta))$ for $j = \{1; 2\}$ and the four different performance measures described in 3.2.3. For this eight sensitivity analysis, a surrogate model was built to approximate the scalar output. The surrogate models based on Polynomial Chaos Expansion (PCE) were constructed using the Least Angle Regression Stagewise (LARS) method to create an adaptive sparse model (Efron et al., 2004). These surrogate models were trained on a dataset comprising 6,000 simulations and validated on an independent set of 2,000 simulations. The accuracy of the metamodels was evaluated on the validation dataset using the predictivity factor Q^2 defined as:

$$Q^2 = 1 - \frac{\sum_{k=1}^N \left(\mathcal{L}(O_j, S_j(\theta^{(k)})) - \hat{\mathcal{L}}(O_j, S_j(\theta^{(k)})) \right)^2}{\sum_{k=1}^N \left(\mathcal{L}(O_j, S_j(\theta^{(k)})) - \frac{1}{N} \sum_{k=1}^N \mathcal{L}(O_j, S_j(\theta^{(k)})) \right)^2}, \quad (16)$$

where $\hat{\mathcal{L}}$ is the surrogate model production, and N is the number of validation samples. For the four performance measures and the two used observations, Q^2 values exceeded 0.9, demonstrating a high level of agreement between the surrogate models and the evaluation of $\mathcal{L}$ for the TELEMAC-2D outputs.

Figures 7 and 8 show the spatial distribution of the total-order Sobol' indices across the 92 subdomains of the study area, for the four performance measures and two observation times. Only a limited number of subdomains exhibit high sensitivity (i.e., high TSI values), indicating that changes in the input parameters are only influential for specific spatial regions. The influential subdomains are consistent across both observations, although some spatial differences emerge between the two acquisition times. For the first observation on February 2, 2021, during the rise of the flood, the regions of high sensitivity are mainly located upstream. In contrast, Figure 8 exhibits more spatial heterogeneity, with additional zones of moderate influence. This can be explained by the propagation of the flood, with more water throughout the study area at the second time of observation, and hence more influential zones downstream than in the first image. Moreover, the larger zones are the most influential since they control most of the flow on the floodplains, and the flooded area is mainly located in these zones.

The dimensionality of the data assimilation (DA) step is reduced by considering only subdomains with total-order Sobol' indices greater than 0.05, highlighted with red contours in Figures 7 and 8. Subdomains that exceed the threshold for at least one of the two observation times (either February 2 or 3) are included in the DA step. In Figure 7, two subdomains have total-order Sobol' indices above the threshold for each of the four performance measures, although the specific subdomains vary depending on the performance measure used. In Figure 8, the number of influential subdomains varies between two and four, depending on the performance measure used. Overall, three subdomains are selected for the DA phase for Cohen's Kappa and the critical success index. In comparison, four subdomains are retained for the Procrustes analysis and the modified Hausdorff distance.

In summary, the sensitivity analysis step significantly reduced the number of parameters for data assimilation, from 92 subdomains to four. In the following, the calibrated subdomains with DA are reported in Figure 9. For Cohen's Kappa and the critical success index, subdomains 1, 2, and 3 are calibrated. For Procrustes analysis and modified Hausdorff distance, subdomains 1, 2, 3, and 4 are calibrated.

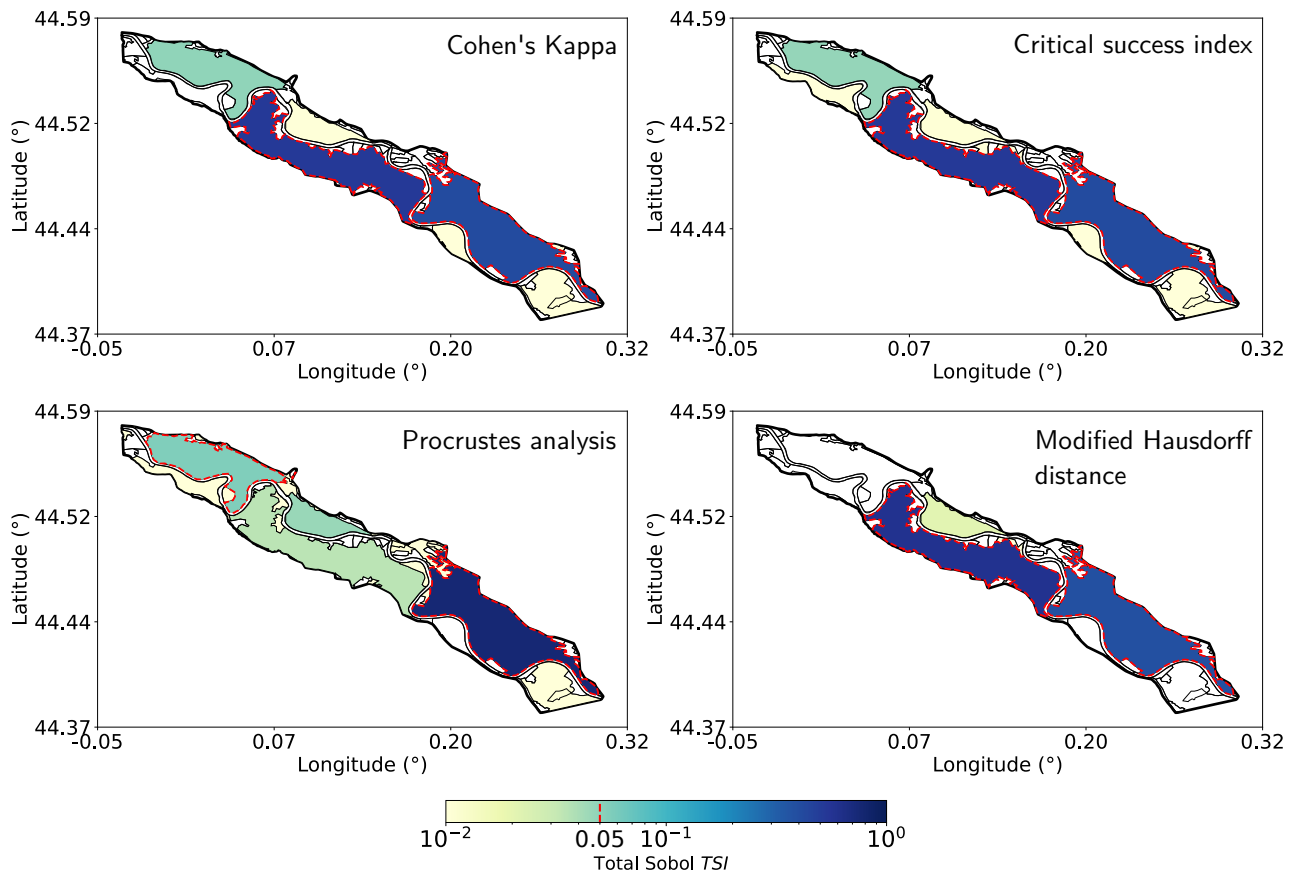


Figure 7: Spatial distribution of total-order Sobol' indices (TSI) across four performance measures using the first reference flood map observation on February 2, 2021, at 17:56 UTC.

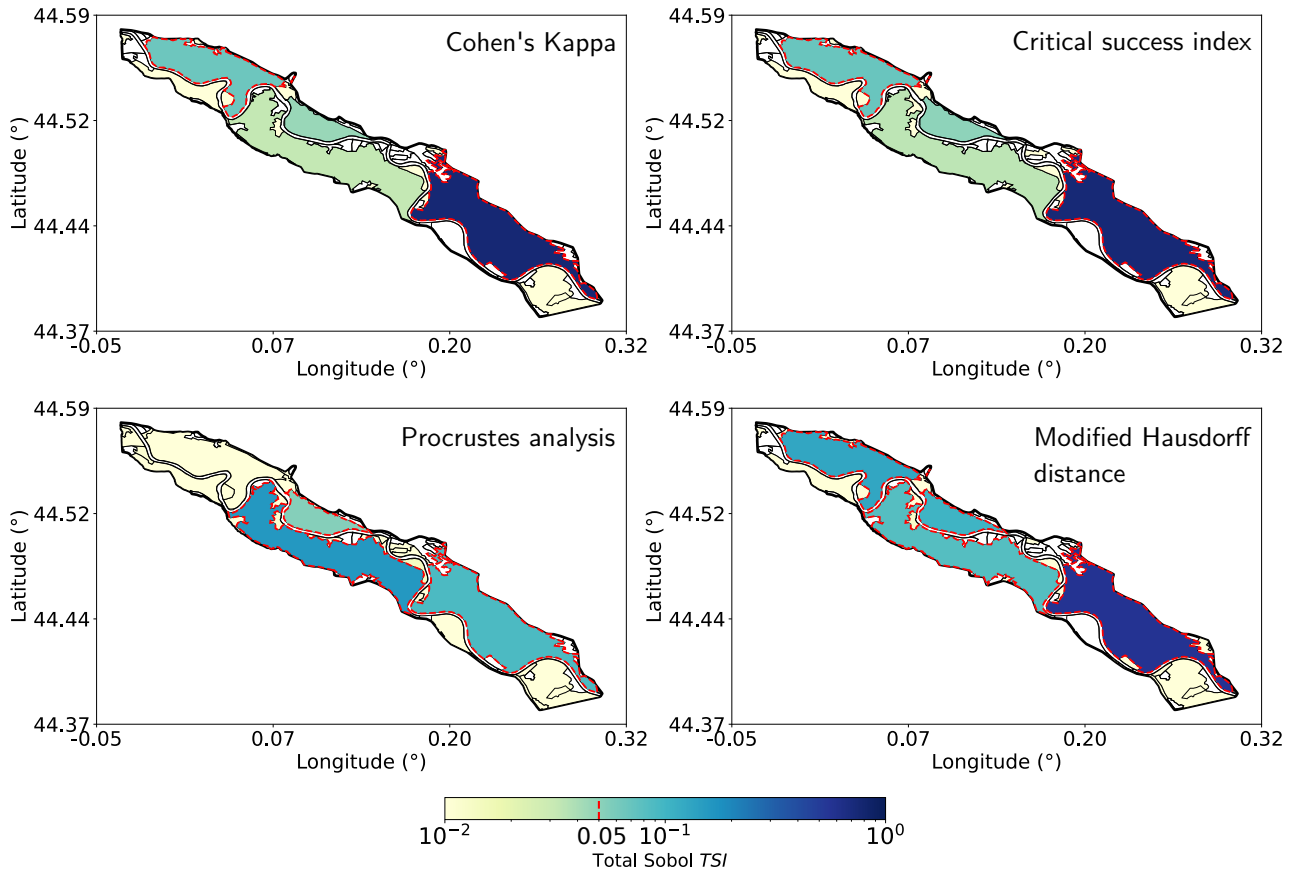


Figure 8: Spatial distribution of total-order Sobol' indices (TSI) across four performance measures using the second reference flood map observation on February 3, 2021, at 17:48 UTC.

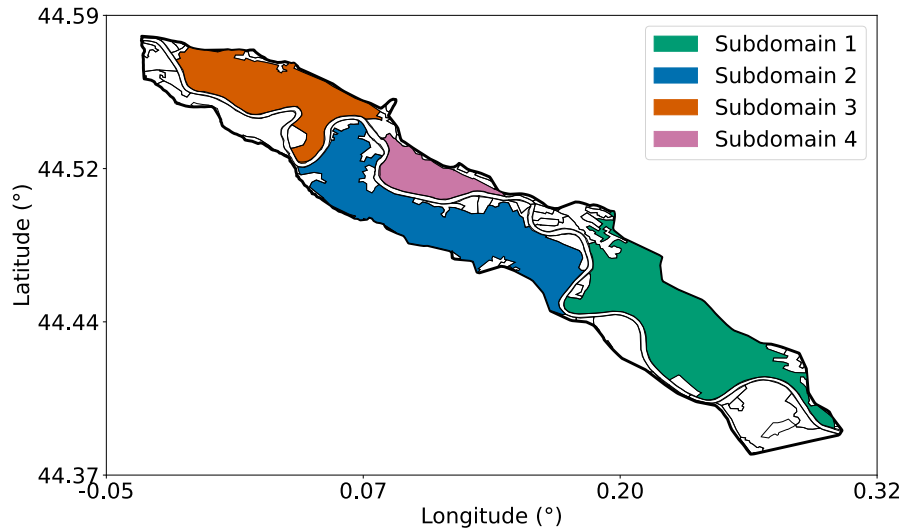


Figure 9: Subdomains 1, 2, and 3 are identified as influential for Cohen's Kappa and the critical success index, while subdomains 1, 2, 3, and 4 are considered influential for Procrustes analysis and the modified Hausdorff distance, based on the sensitivity analysis.

5 Results of the data assimilation step

The influential parameters identified in Section 4 are calibrated using the 3D-Var algorithm for both the twin and real experiments, following the methodology described in Section 3.2. Throughout the data assimilation (DA) step, the background error covariance matrix $\mathbf{B}$ is fixed to $\mathbf{B} = 100\mathbf{I}_2$ with $\mathbf{I}_2$ the identity matrix of size 2, based on the uncertainties associated to roughness parameters described in the literature (standard deviation of the Strickler coefficient is fixed to $10 \text{ m}^{1/3}/\text{s}$).

5.1 Twin experiment

In the twin experiment, the synthetic observations were assumed to be noise-free and perfectly accurate by setting $\omega = 10^{-10}$, thereby assigning a very high confidence to the observation term in the cost function. The influence of the choice of performance measure on the calibration of the Strickler coefficients was evaluated for the four performance measures: Cohen's Kappa, critical success index, Procrustes analysis, and modified Hausdorff distance. For each calibrated subdomain, the initial Strickler coefficient guess θ^b was randomly drawn from a uniform distribution $\mathcal{U}(5, 50)$ to consider a broader uncertainty than usual flood modeling studies as defined in Table 1. Since 3D-Var can be sensitive to the initial parameter guess due to the shape of the cost function, 10 assimilations were performed per performance measure for 10 random initializations sampled in a uniform distribution for each parameter $\mathcal{U}(5, 50)$. In the 3D-Var minimization, the parameters are constrained between 5 and $50 \text{ m}^{1/3}/\text{s}$.

Figures 10 and 11 show the evolution of the Strickler coefficient estimates over iterations in subdomains 1 and 3, respectively, for each of the four performance measures. In subdomain 1 (see Figure 10), located upstream in the domain, all four performance measures lead to convergence of the 3D-Var estimates toward the true Strickler value (dashed gray line). For Cohen's Kappa and CSI, the convergence to the true values is fast with the ensemble of 10 runs converging around the true value (12.5) within 4 to 5 iterations. In contrast, Procrustes analysis and modified Hausdorff distance show slower convergence and greater variability in the initial iterations, but stabilize around the true value after approximately 6 to 7 iterations.

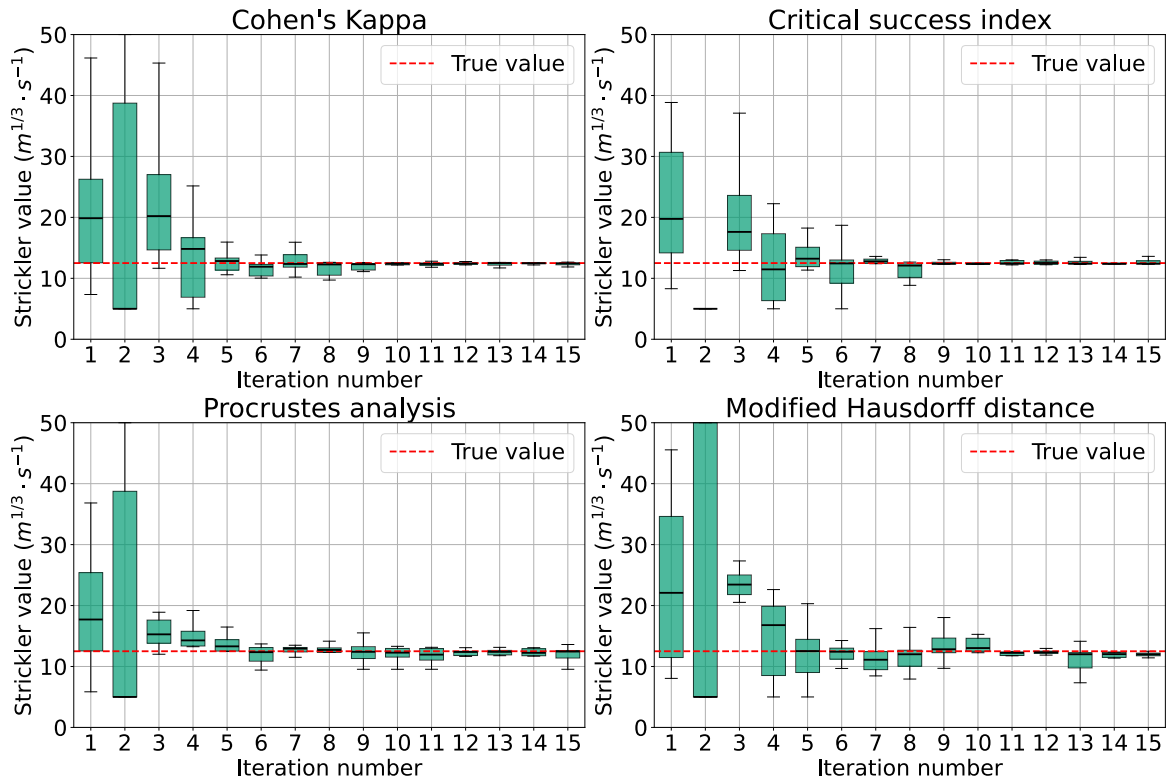


Figure 10: Estimates of the Strickler values for the twin experiment in subdomain 1 for four performance measures using 3D-Var. Boxplots show the variability of estimates across 10 initializations over iterations.

In subdomain 3 (see Figure 11), convergence to the true values is again observed for Cohen's Kappa, CSI, and modified Hausdorff distance. However, CSI shows greater variability in the first iterations compared to Cohen's Kappa. When

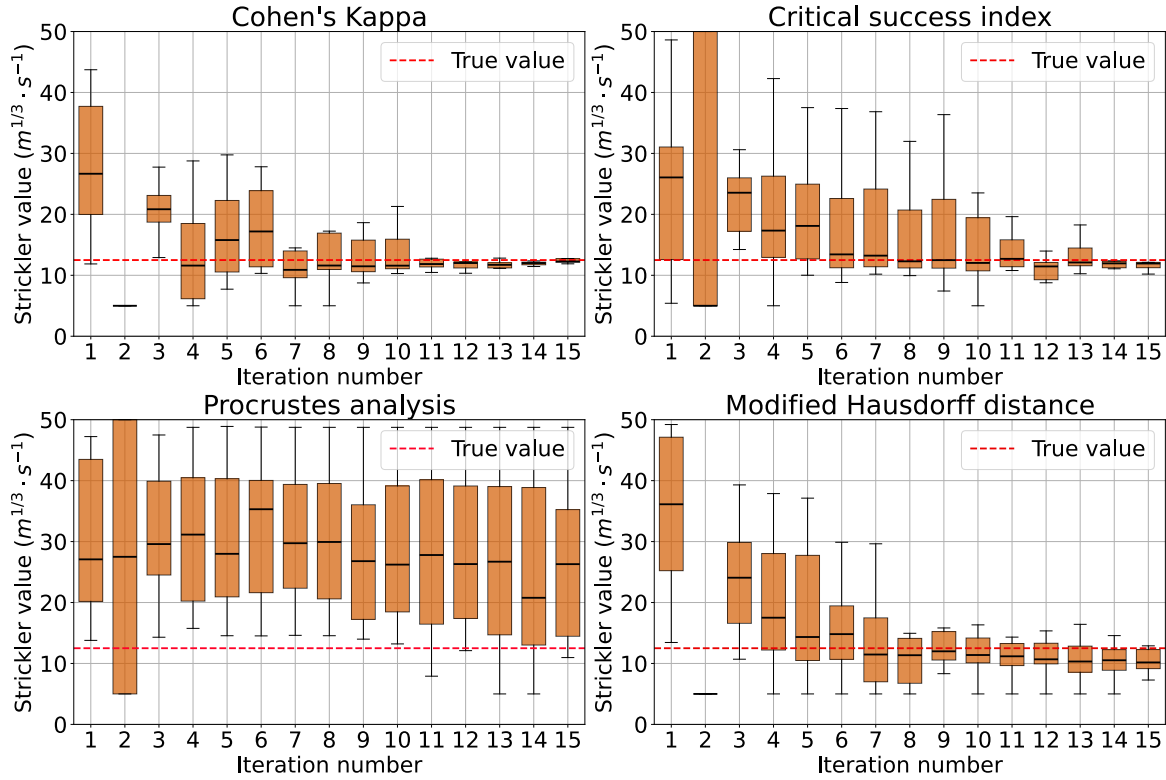


Figure 11: Estimates of the Strickler values for the twin experiment in subdomain 3 for four performance measures using 3D-Var. Boxplots show the variability of estimates across 10 initializations over iterations.

using Procrustes analysis, 3D-Var fails to converge to the true value. In Travert et al. (2025a), the Procrustes analysis was not considered as a reliable performance measure because it was not equally sensitive to an increase or a decrease in the flood extent size. It might be the cause of the inability to reach the true value in subdomains 2, 3, and 4. Furthermore, the lower influence of these subdomains on the Procrustes analysis performance measure is highlighted in Figure 8. For the first observation, the total Sobol' index in subdomain 2 is higher for the other performance measures, and similarly for subdomain 3 for the second observation.

Results for subdomains 2 and 4 are presented in Figures 23 and 24 in Appendix A, showing behavior similar to that observed in subdomain 3. When using Procrustes analysis as a performance measure, the Strickler values are consistently overestimated in subdomains 2, 3, and 4. Consequently, applying this performance measure for flood map assimilation results in underestimation and overestimation of water depths. Figure 12 (a) shows the water depth differences between the simulation calibrated using Procrustes analysis (with mean Strickler values from the final iteration) and the true state generated with a reference Strickler field. On February 2, 2021, the results indicate an overestimation of water depth downstream and an underestimation in the central part of the study area (ranging from 10 to 20 cm), attributed to discrepancies in Strickler values. On the contrary, the same analysis is carried out in Figure 12 for Cohen's Kappa formula and shows that the differences in water depth are negligible due to Strickler values close to the reference values after the calibration process.

After considering perfect observations, we generated synthetic noised observations using the method of characteristics for the advection scheme. The observation error ω was quantified by evaluating the performance measure between the noised observations and the perfect observations. For instance, Figure 13 shows the Strickler estimates on subdomains 1 and 3 when assimilating this uncertain observation using 3D-Var with Cohen's Kappa performance measure. The same prior as the one used for perfect observations assimilation were used. The Figure suggests that the assimilation of noised synthetic observations allows to find a compromise between the prior and the true reference value. Indeed, the Strickler estimates are closer to the reference than prior to assimilation even when assimilating uncertain observations. For the other performance measures, the behavior was identical. In the future, more noise levels should be tested to find up to which noise levels, uncertain satellite flood maps provide information for calibrating the flood model.

Overall, in the twin experiment configuration, the 3D-Var algorithm converges more efficiently and reliably when using Cohen's Kappa and modified Hausdorff distance as performance measures. These findings are consistent with those of

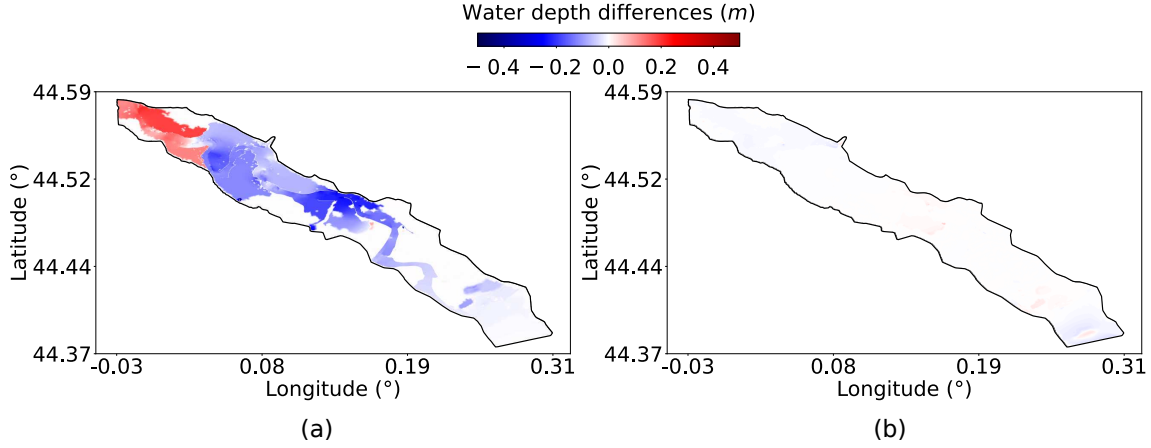


Figure 12: Water depth differences between simulations calibrated with 3D-Var (mean final estimates of the 10 runs) using (a) Procrustes analysis or (b) Cohen's Kappa and the true state on February 2, 2021, at 17:56 UTC.

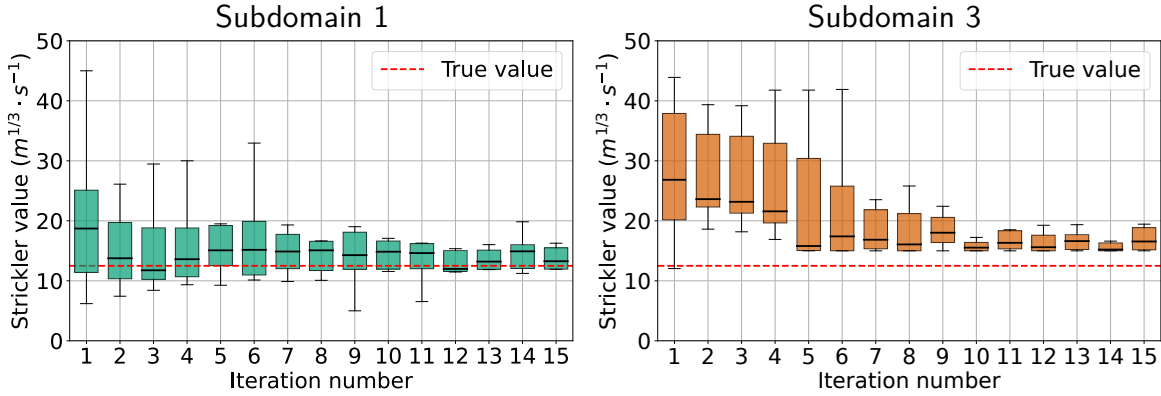


Figure 13: Estimates of the Strickler values for the twin experiment in subdomains 1 and 3 for Cohen's Kappa formula using 3D-Var. Boxplots show the variability of estimates across 10 initializations over iterations.

Travert et al. (2025a), who identified Cohen's Kappa and modified Hausdorff distance as robust metrics for comparing simulated and observed flood maps. In contrast, CSI and Procrustes analysis showed less reliability for estimating the parameters in the subdomains.

5.2 Real experiment with observations from satellite imagery

In this section, the results of the data assimilation using flood maps derived from SAR imagery are analyzed. Figures 14 and 15 show the estimated Strickler values for subdomains 1 and 3, respectively. The values of B and of the observations weight ω used in the twin experiment are conserved assuming that the observations are perfect.

In subdomain 1 (see Figure 14), all four performance measures show consistent convergence. Cohen's Kappa, CSI, and modified Hausdorff distance converge rapidly with reduced spread in the estimates after five iterations. Procrustes analysis shows higher variability, even though the mean estimate converges to approximately $33 \text{ m}^{1/3}/\text{s}$. However, for all performance measures, the Strickler values consistently reach high values, which is unlikely in floodplains where roughness is usually high (corresponding to low Strickler values).

Subdomain 3 (see Figure 15) presents a different pattern. For Cohen's Kappa and the modified Hausdorff distance, the estimated Strickler values rapidly converge toward the upper bound of the defined parameter space, indicating that the assimilation compensates for observation discrepancies with the model by maximizing the roughness parameter. In contrast, the estimates based on the critical success index and Procrustes analysis show more variability. The fact that the estimates often reach high Strickler values in the floodplains or remain at the bounds of the parameter range suggests that the optimization is constrained by the limits of the parameter space. Furthermore, at the end of the assimilation process, the value of the cost function is still high and not close to zero, suggesting that the estimated Strickler still

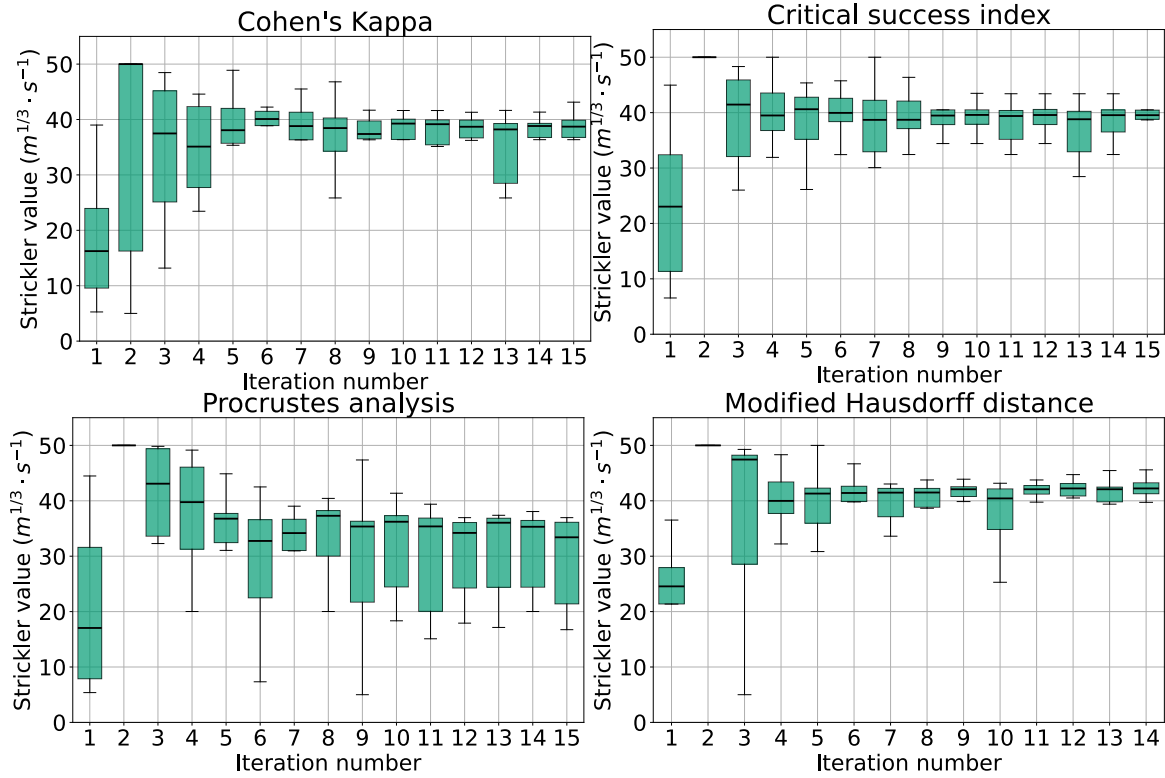


Figure 14: Estimates of the Strickler values for the real experiment in subdomain 1 for four performance measures using 3D-Var. The boxplots show the variability of estimates across iterations for 10 different initializations.

leads to important differences between the simulated and observed flood maps. The simulated flood extent may be over- or underestimated due to uncertainties not taken into account (e.g., inlet discharge, dike modeling), or the derived observations from satellite images may be too uncertain. For instance, in Travert et al. (2025b), it was shown that the variability of the flood maps highly depends on the processing choices used to treat the satellite image.

In Figure 16, the observed flood maps at the two times of observation are compared to the simulated flood maps at the end of the assimilation for Cohen's Kappa performance measure. After the calibration, the simulated flood map and the observed flood map have the same geometric structure, but differ slightly especially for the 2nd of February 2021 in the center and downstream of the model, with over and underestimation of the flooded areas. Upstream of the domain, the wave front is not perfectly aligned with the simulated front going faster than the observed front. These differences could be due to a wrong parameterization of the model, because the uncertain observations were considered perfect in this first approach, or because the model is not able to reproduce the flood dynamics due to missing flood processes or too constrained model with fixed minor bed for instance.

In Section 5.2.1, flood maps generated with other flood mapping methods are assimilated considering perfect observations to show the variability due to the choice of processing method of the observations. Then, Section 5.2.2 explores the role of the weighting factor ω in the cost function to account for the observation errors. Finally, Section 5.2.3 presents the assimilation of the flood maps with a more uncertain flood model, considering the channel roughness as an uncertain parameter.

5.2.1 Impact of observation selection

In the first assimilation procedure, the assimilated observations were generated using a local thresholding method. However, the generated flood maps can be imperfect due to measurement errors, the presence of vegetation, or the choice of processing methods (Landuyt et al., 2018; Zhao et al., 2021; Travert et al., 2025b).

The impact of these uncertainties is evaluated by assimilating flood maps derived with another flood mapping method. A pre-trained Convolutional Neural Network (CNN), trained to identify flooded regions on Sentinel-1 images (Bonafilia et al., 2020), was used to generate the flood maps. Figure 17 presents the resulting flood maps, which appear smoother, with fewer holes and larger floodplain coverage, compared to the flood maps generated using the local thresholding

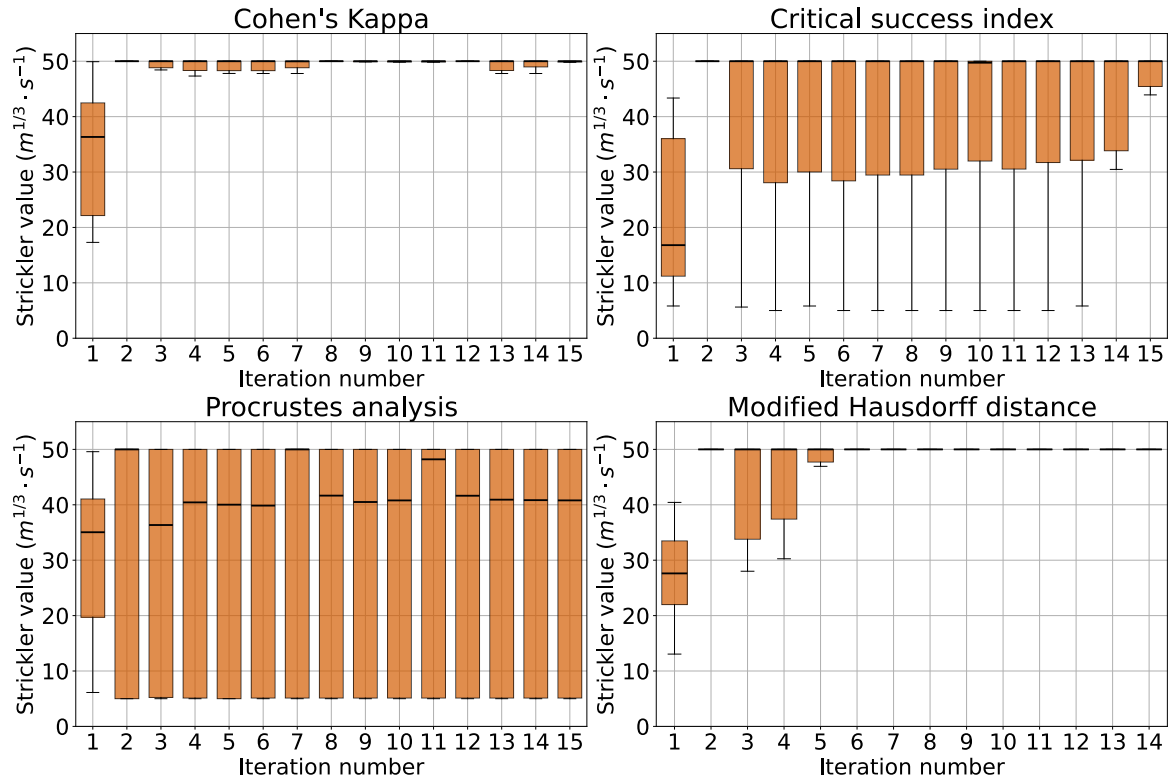


Figure 15: Estimates of the Strickler values for the real experiment in subdomain 3 for four performance measures using 3D-Var. The boxplots show the variability of estimates across iterations for 10 different initializations.

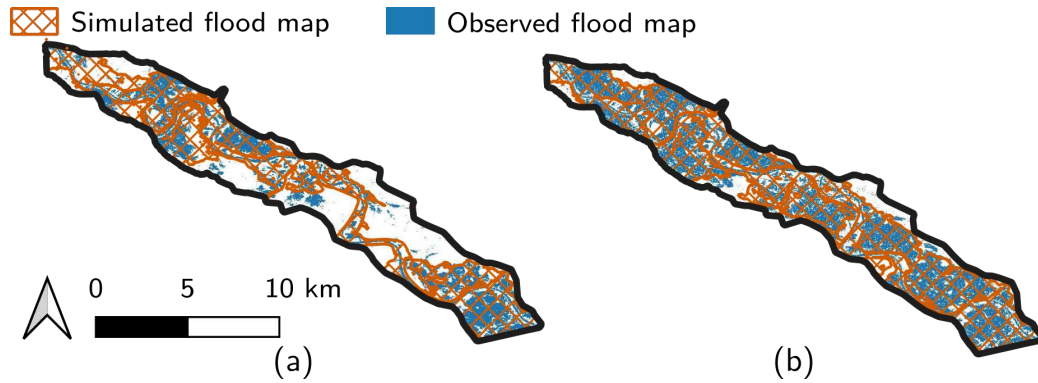


Figure 16: Comparison of the observed flood maps on (a) February 2, 2021, at 17:56 UTC, (b) February 3, 2021, at 17:48 UTC, with the simulated flood maps at the end of the assimilation for one of the assimilation with Cohen's Kappa performance measure.

method (see Figure 4). For instance, the variation in the flooded area between the two flood mapping methods reaches 10 km^2 for the flood map generated on February 2.

Figure 18 compares the Strickler values estimated using Cohen's Kappa performance measure for flood maps derived from local thresholding and CNN methods, using the same initial guess. In subdomain 1, the same tendency is identified, with high values of Strickler for assimilating either observations derived from CNN or local thresholding methods. However, for subdomain 3, the estimated Strickler converges either to the upper bound or to the lower bound, depending on the observations. Multiple parameter combinations may fit the observations equally well (equifinality issues). However, the cost functions when assimilating local thresholding and CNN-based flood maps do not converge to the same values, suggesting that it is not an equifinality problem, and it is due to the variations in the generated flood maps.

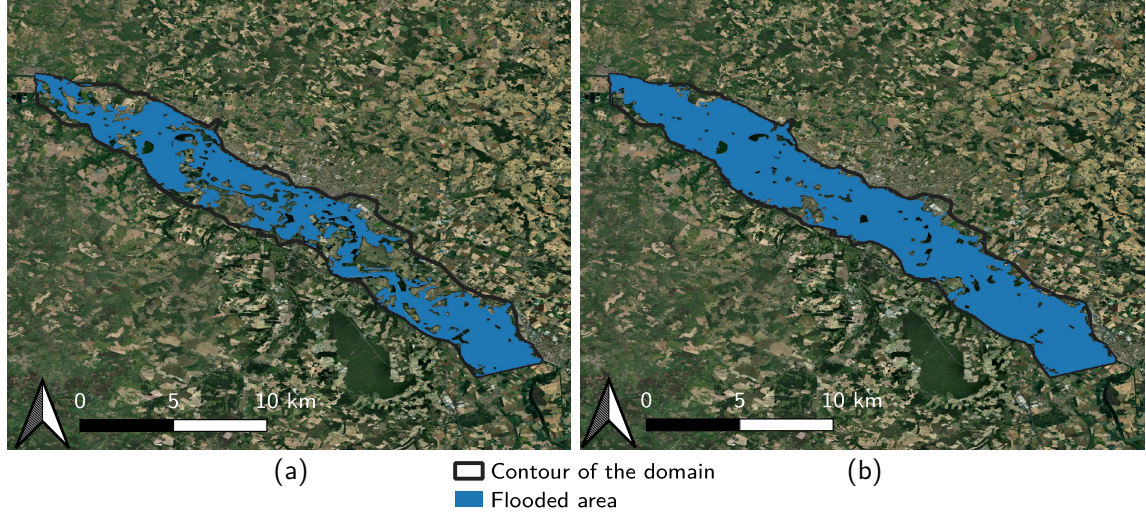


Figure 17: Flood maps extracted from Sentinel-1 acquisitions using Convolutional Neural Networks on the Garonne River on (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

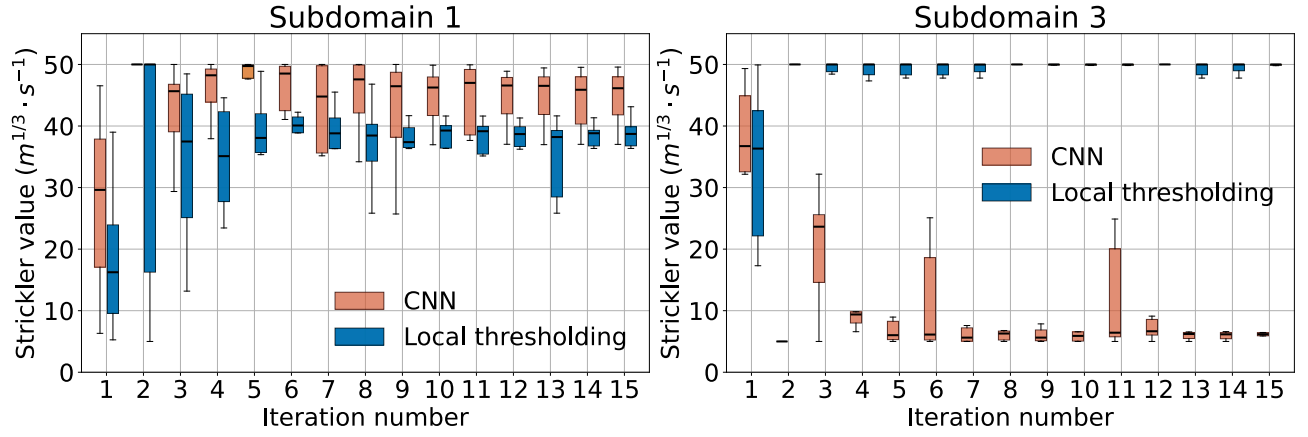


Figure 18: Estimates of the Strickler values in subdomain 1 and 3 using Cohen's Kappa performance measure in the 3D-Var cost function. The blue and red boxplots represent the variability of the estimations for the assimilation of flood maps derived using local thresholding and Convolutional Neural Networks (CNN), respectively.

These results highlight that due to the uncertainty in the satellite-derived observations, the distributed Strickler are difficult to characterize. A more appropriate approach would involve adjusting ω , allowing for a more realistic weighting between the background term and the observation term in the 3D-Var cost function to account for observation uncertainties. However, determining an optimal weight between J_b and J_o is complex, as the observations are categorical flood maps (matrices of 0 and 1) and the performance measures used are non-Euclidean, making it difficult to define statistically covariance matrices for observation errors. A sensitivity analysis of the assimilation results with respect to variations in ω is presented in Section 5.2.2. The aim is to evaluate the influence of the uncertainty of the observed flood maps on the calibrated roughness field.

5.2.2 Impact of the weighting factor in the cost function

To account for the uncertainties in the generation of flood maps from satellite observations, the weighting factor ω in the cost function should be modified. We aim to balance the empirical relative variation (i.e., standard deviation) of the background and observation terms in the cost function defined in Equations (9). Rather than arbitrarily selecting ω , a target ratio $\gamma \in]0, 1[$ is defined representing the desired ratio between the standard deviations of the background and observation cost terms across a range of parameter values:

$$\frac{\text{Std}(J_b(\theta))}{\text{Std}(J_o(\theta))} = \frac{\gamma}{1 - \gamma}, \quad (17)$$

where Std is the standard deviation. For instance, for $\gamma = 0.5$, the variations of the background and observation terms are balanced, meaning that the total cost function variation due to θ should give equal importance to the background and observation terms.

The matrix $\mathbf{B}$ was considered as a scalar multiple of the identity (no correlated errors), such that:

$$\mathbf{B} = (\sigma^b)^2 \mathbf{I}. \quad (18)$$

Then, using Equations (17) and (18):

$$\frac{1}{(\sigma^b)^2 \omega} = \frac{\gamma}{1 - \gamma} \cdot \frac{\text{Std}(\mathcal{L}(O_1, S_1(\theta))^2 + \mathcal{L}(O_2, S_2(\theta))^2)}{\text{Std}(\|d_b\|^2)}, \quad (19)$$

where $d_b = (\theta - \theta^b)$. The standard deviations are estimated empirically using 500 samples of $\theta \sim \mathcal{U}(5, 50)$. The background state is set to $\theta^b = [12.5; 12.5; 12.5]$ for CSI and κ , and to $\theta^b = [12.5; 12.5; 12.5; 12.5]$ for Modified Hausdorff Distance and Procrustes Analysis, corresponding to the barycenter (for croplands with tall vegetation) of the ranges described in Table 1. In this study, $\gamma \in \{0.01; 0.1; 0.2; \dots; 0.8; 0.9; 0.99\}$ to evaluate the sensitivity of the calibration to the relative trust placed in the background versus observation terms. A low γ assumes that the satellite-derived flood maps, while uncertain, are the most certain source of information, and inversely for high γ .

We highlighted in both the twin experiment and the first assimilation of the real experiment that the critical success index and Procrustes analysis were not adapted for the assimilation of flood maps in this test case. Thus, only the results for Cohen's Kappa and modified Hausdorff distance are reported. The estimated weights ω for the range of selected γ values for a fixed background covariance matrix $\mathbf{B} = 100\mathbf{I}_2$ are reported in Table 2. Given the values of ω for the different levels of confidence between the background and misfit to observations terms, the observations introduced in Section 2.4.2 and reported in Figure 4 are assimilated.

Table 2: Computed ω for each γ value for Cohen's Kappa and modified Hausdorff distance, with fixed $B = 100\mathbf{I}_2$.

γ	ω (Cohen's Kappa)	ω (modified Hausdorff distance)
0.01	1.02e+01	5.05e-03
0.10	1.12e+02	5.56e-02
0.20	2.53e+02	1.25e-01
0.30	4.33e+02	2.15e-01
0.40	6.71e+02	3.34e-01
0.50	1.01e+03	5.00e-01
0.60	1.52e+03	7.52e-01
0.70	2.36e+03	1.17e+00
0.80	4.03e+03	2.00e+00
0.90	9.09e+03	4.50e+00
0.99	1.00e+05	4.95e+01

Figures 19 present the variation of the floodplain Strickler values per iteration of 3D-Var for the γ values reported in Table 2 using Cohen's Kappa formula and modified Hausdorff distance in the cost function, respectively. The same initial guess θ^b is used for all γ values.

For the two performance measures, the sensitivity to γ is similar for subdomains 1 and 3. For both subdomains, low values of $\gamma \leq 0.2$, corresponding to low confidence in the observations, lead to Strickler values converging near the initial guess. As γ increases (more trust placed in the observations), the optimization becomes more sensitive to variations around the initial guess. The Strickler values converge to values near the initial guess except for high confidence in the observations. For high γ values, the cost function minimization mainly relies on the minimization of the errors between the observed and simulated flood maps, with little effect of the background knowledge. In that case, if the observation is very different from the simulations, the Strickler values can vary significantly, as seen in Figure 19 for $\gamma = 0.99$. The cost function, is quite smooth, but as γ increases some local minimum appear, and the cost function minimization process can get stuck in them. For instance, Figure 20 shows the cost function shape for Cohen's Kappa performance for coupled variations of the Strickler coefficients. While for $\gamma = 0.1$ there seems to be a global minima even if the visualization is only for coupled variations two by two, while the cost function is less smooth for $\gamma = 0.8$. For $\gamma = 0.8$, Figure 21 shows the estimated Strickler with 3D-Var in subdomain 3 for several values of θ^b by sampling its probability density distribution. For this observation noise level the cost function is difficult to minimize and get stuck in the local minimum shown in Figure 20. The optimization procedure gets stuck in three zones. The experience modeler could choose the appropriate Strickler value among the three zones considering the land use and the generated

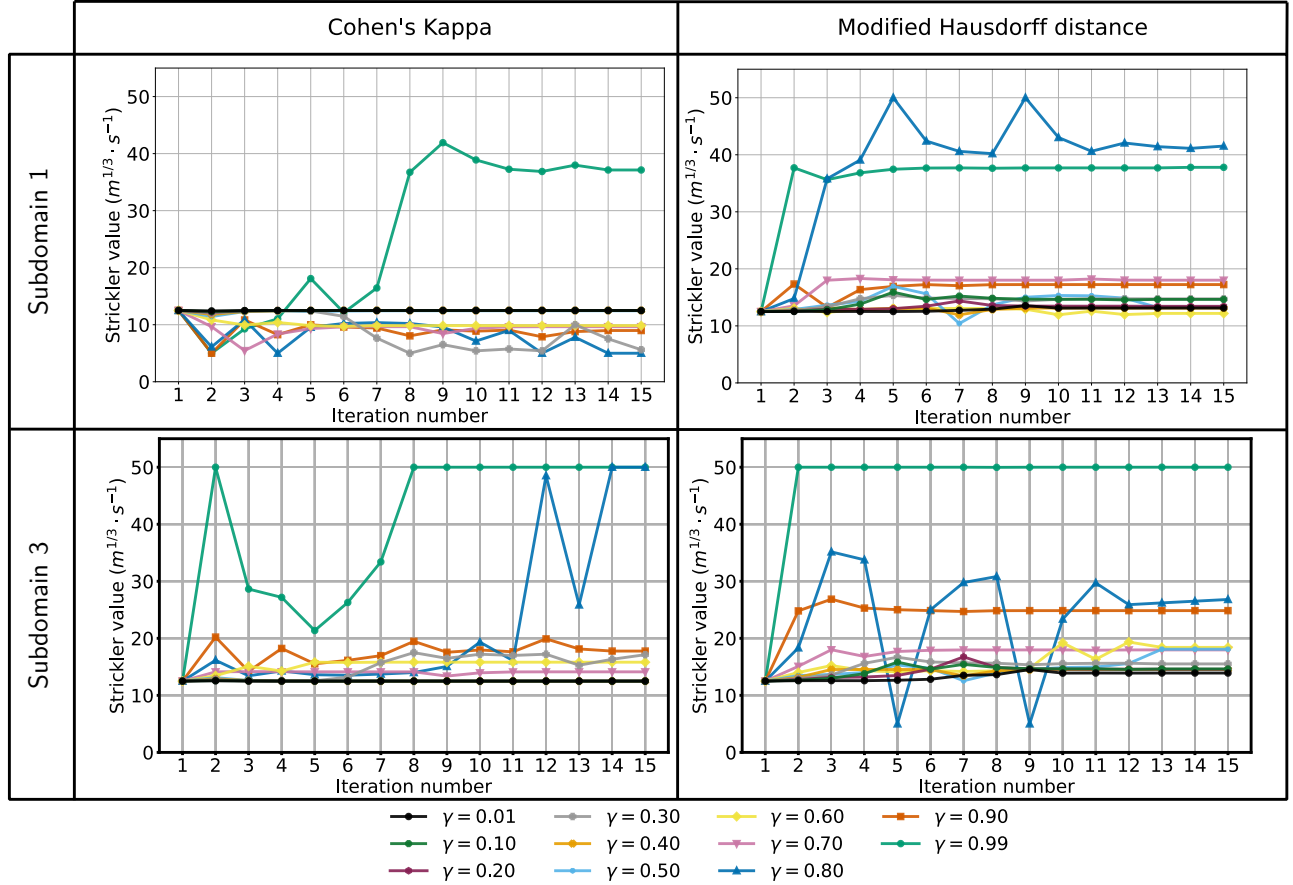


Figure 19: Estimates of the Strickler values for the real experiment in subdomains 1 and 3 for two performance measures using 3D-Var and varying confidence (γ) in the observation term.

flood maps. Another possibility is to use other algorithms such as stochastic gradient descent or genetic algorithms to find a global minimum for this level of confidence in the observations, but it would require a higher computational time.

This analysis underscores the importance of carefully tuning the weighting factor ω in 3D-Var frameworks for floodplain calibration and check that the global minimum of the cost function is found. Depending on the level of confidence in the observations and the background, the Strickler estimates in the floodplains can vary significantly. The specific value of γ to choose for flood maps assimilation is still an open question. The results suggest that improvements on the flood mapping procedure should be made to reduce the uncertainties for floodplain roughness calibration. Furthermore, the weight factor should be quantified more accurately to estimate where the Strickler values should lie when combining background knowledge and observations.

While γ is difficult to quantify for SAR derived flood maps and for specific performance measures, another possibility is to consider that the observations are perfect and reduce the constraint on the models by choosing more uncertain parameters. This analysis is carried out in Section 5.2.3.

5.2.3 Accounting for other uncertainties than floodplain roughness

In this 2D flood model, initially, only the roughness of the floodplains was assumed uncertain, with fixed channel roughness based on previous studies and calibration using in-situ data. However, as discussed in Section 5, the convergence of Strickler values in subdomain 3 to the bounds of the parameter range indicated that other sources of uncertainty may be unaccounted for (if considering that the observations are perfect and noise-free), limiting the model's ability to reproduce the observed flood maps.

Then, model constraints were relaxed by allowing the channel roughness to vary. Instead of using fixed Strickler values in the riverbed, the three roughness subdomains are considered bounded between 30 and 50 $m^{1/3}/s$. With this updated uncertainty representation, the data assimilation framework was applied for one random initial guess of the parameters,

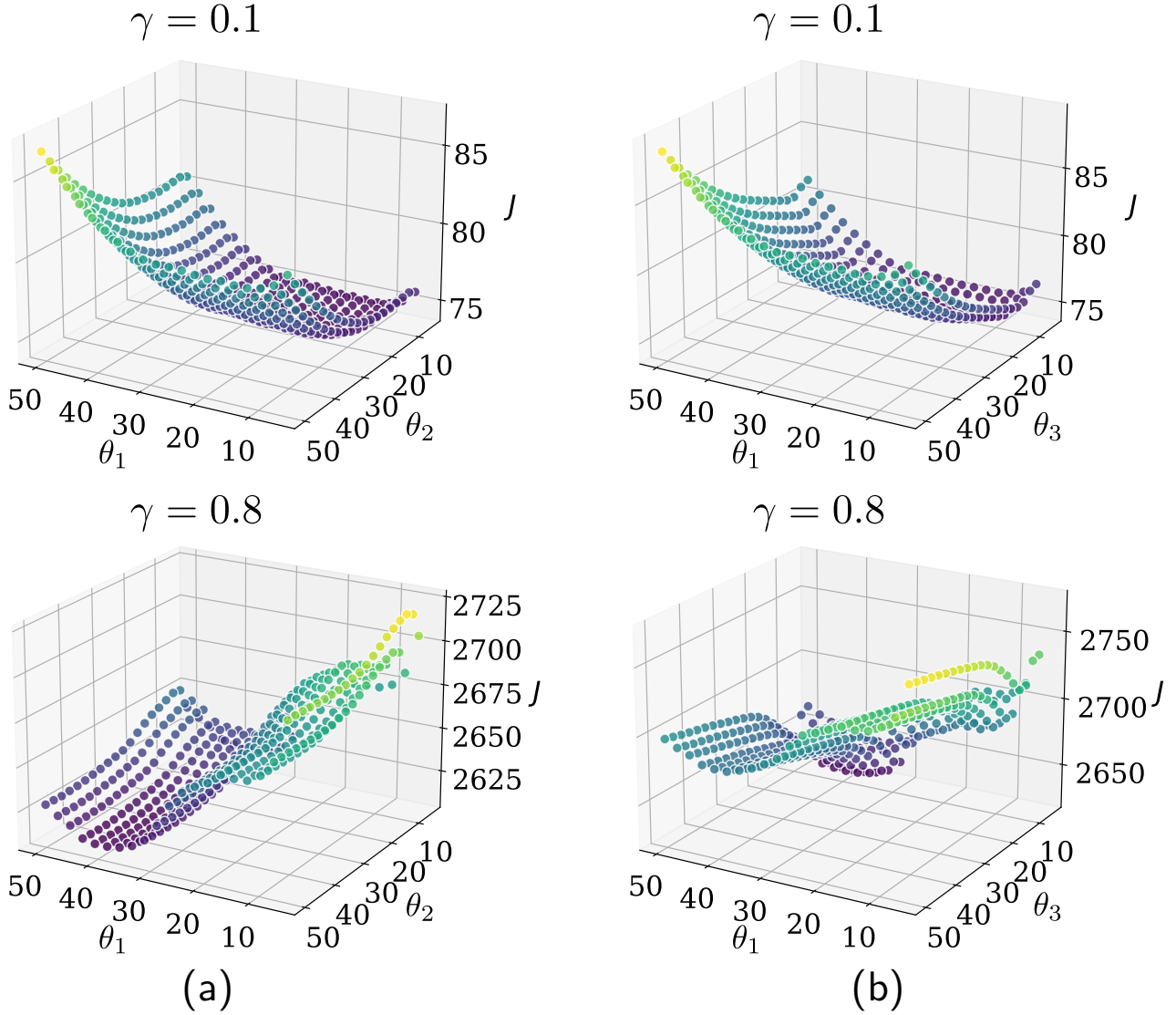


Figure 20: Variation of the cost function J for Cohen's Kappa performance measure $\mathcal{L}$ for $\gamma \in \{0.1; 0.8\}$ as a function of the coupled parameters (a) θ_1 and θ_2 for $\theta_3 = 12.5$, and (b) θ_1 and θ_3 for $\theta_2 = 12.5$.

assimilating observations derived from both the local thresholding and CNN methods on SAR imagery. Figure 22 shows the Strickler evolution as a function of the number of algorithm iterations for the influential subdomains (subdomains 1, 2, and 3) in the floodplains and the three channel subdomains (Tonneins/Mas d'Agenais, Mas d'Agenais/Marmande, and Marmande/La Réole). Two set of observations are tested by assimilating flood maps derived from two different SAR classification methods: local thresholding (solid lines) and a Convolutional Neural Network (CNN) (dashed lines). The cost function gradients were primarily driven by the channel parameters rather than the floodplains. This illustrates the dominant hydraulic control exerted by the channel during overbank flows, which governs not only in-channel velocities but also the timing and extent of floodplain inundation. With that configuration, the Strickler values do not converge to the bounds of the parameter range, suggesting that the model can reproduce the observations better within the parameter range. During the initial iterations, most subdomains have high adjustments in their Strickler values, but then only the channel roughness keeps evolving up to 10 iterations.

In the riverbed subdomain and floodplain subdomains, the assimilation using flood maps derived from local thresholding does not converge to the same Strickler values compared to the assimilation using flood maps derived with CNN. It underscores the need for better handling of observation uncertainty, as the results are still conditioned on the assumption of perfect observations, which is rarely realistic in practice.

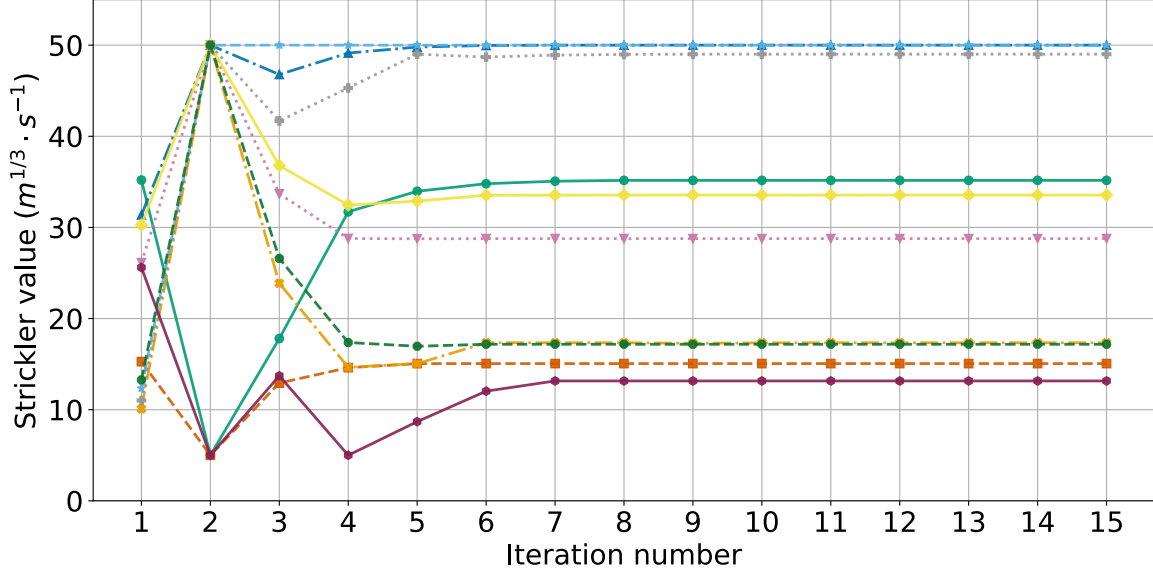


Figure 21: Estimates of the Strickler values for the real experiment in subdomain 3 for Cohen’s Kappa performance measure using 3D-Var with $\gamma = 0.8$ and different θ^b .

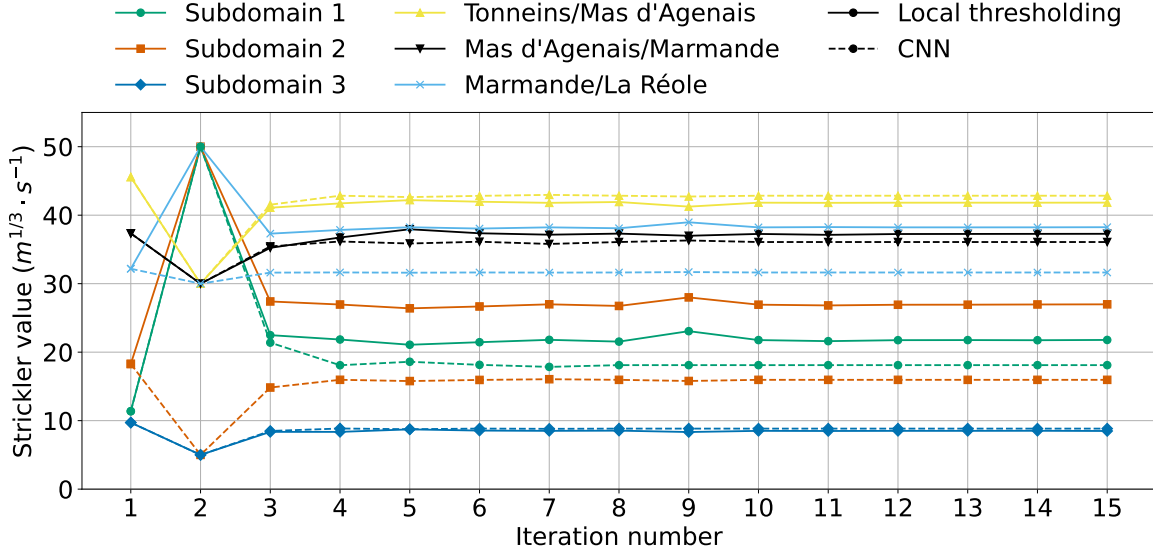


Figure 22: Estimated Strickler across iterations of 3D-Var with Cohen’s Kappa performance measure for assimilating flood maps derived from local thresholding (solid lines) or Convolutional Neural Network (dashed lines) methods.

6 Discussion

This study explored the assimilation of flood maps from SAR imagery into a 2D flood model using a 3D-Var variational approach to calibrate spatially distributed roughness parameters. While the results are promising on controlled experiments, several limitations impact the robustness of the calibration in real-world applications.

The main challenge in assimilating flood maps with the presented performance measures is to include observation uncertainty into the 3D-Var cost function via the weighting factor ω . The uncertainty in the SAR-derived flood maps was only implicitly addressed by tuning the relative weights between the background term and the observation term instead of using known uncertainties of the observations. This approach poorly reflects the true uncertainty in SAR-derived flood maps, which can be affected by measurement errors, speckle noise, or segmentation errors (Zhao et al., 2021; Landuyt et al., 2018). Future work should focus on more robust estimation of ω , especially when the misfit between

the observed and simulated flood maps is quantified using non-Euclidean performance measures. Inflation techniques (Carrassi et al., 2018) by scaling or tuning $\mathbf{B}$ or ω could help manage the trade-off between trust in the observations and the background term, especially when satellite-derived flood maps are noisy or inconsistent. Another possibility is to empirically estimate the observation errors of the performance measures evaluation. For example, flood maps can be generated with many flood mapping algorithms and hyperparameter settings (Travert et al., 2025b). Then, for all these flood maps, the distribution of the performance measures can be computed to provide an empirical estimate of the uncertainty. However, these estimates should be presented with a probabilistic framework, assigning likelihoods to each generated flood map, as the results of some processing methods lead to flood maps closer to reality than others. This approach only considers the uncertainty introduced during the SAR image processing. However, SAR imagery may be affected by measurement errors, geolocation inaccuracies, incidence angle effects, etc. These errors on the raw SAR images should also be accounted for in the generation of the flood maps. In this study, the twin experiment was carried out only considering perfect observations, and one level of observation noise. In the future, more noise levels should be tested with the synthetic observations to evaluate in more detail the role of the weighting factor and the possibility to calibrate floodplain roughness. However, adding synthetic noise to the synthetic observations consisting of binary flood maps while simultaneously evaluating the weighting factor for this level of noise is an open question and was not explored in this study.

Another limitation in this study is that the flood maps can be significantly different from the simulated ones. This suggests that the parameter space is too narrow to match the observations, and that the assimilated observations are outside the model's solution space. However, as long as the observation uncertainties are not clearly defined, it is complicated to state whether the parameter range is too narrow or if the differences are only explained by the uncertainties. In the cost function, the model error is not accounted for, and we only considered that the variability of the model outputs is ruled by its parameterization. In future studies, the model error and missing physical processing could be added with an additional term in the cost function. Then, we could define a distribution of probability and a confidence region around each optimal Strickler value by sampling noise model errors. This model error could be introduced on the velocity field through a Gaussian noise, but this is difficult due to the use of different performance measures and of the computational cost of launching several DA when sampling the model error distribution.

Then, when assimilating flood maps, the modeler should question if the flood maps are reliable enough to calibrate a set of parameters or if the flood maps can be used only to calibrate a less constrained model with additional uncertain parameters (e.g., here the channel roughness) or wider parameter ranges.

7 Conclusion and perspectives

This study investigated the assimilation of uncertain flood maps derived from satellite observations to estimate a spatially distributed roughness field in a 2D flood model using variational data assimilation (3D-Var). The objective was to evaluate how different performance measures quantifying the misfit between the simulated and observed flood maps influenced the calibration of the Strickler field in the assimilation process.

Using a two-phase strategy that combines sensitivity analysis and variational assimilation, we first demonstrated that only a subset of roughness parameters is influential in evaluating the performance measures. The sensitivity analysis reduced the number of parameters from more than 90 to 3 or 4 floodplain roughness subdomains, depending on the performance measures. In controlled twin experiments using synthetic observations, the assimilation successfully recovered the true parameter values when using Cohen's Kappa and the modified Hausdorff distance, assuming perfect observations. However, for the two other performance measures, Procrustes analysis and critical success index, the assimilation did not converge to the true values or exhibited more variability depending on the initial guess, indicating the influence of the choice of performance measures in the 3D-Var assimilation. When applied to real flood events using SAR-derived flood maps, the results confirmed the impact of the performance measure. The Cohen's Kappa and modified Hausdorff distance resulted in converged results in contrast to the two other performance measures. However, the results also indicated that calibrating only floodplain roughness is insufficient when real flood maps are too uncertain and are considered noise-free in the 3D-Var cost function. The inclusion of channel roughness as an additional uncertain parameter improved the model sensitivity and prevented convergence to the parameter bounds. The second option is to weight the 3D-Var cost function by considering the uncertainties of the background and observation terms. We conducted a sensitivity analysis of the degree of uncertainty of the observation term in the total cost function. We highlighted the variations of the calibrated parameter depending on the weight of the relative variation of the observation term compared to the background term. In the future, additional experiments on the twin experiment with noise on the synthetic observations should be conducted to determine the maximum level of observation noise that provide information for calibrating floodplain roughness using uncertain flood maps. For each assimilation, in the future, we could also introduce a model error to account for unmodeled physical processes or bad flood dynamics representation, to define a region of confidence around each optimal state determined using 3D-Var.

To improve flood model calibration and the assimilation of flood maps, future work should focus on quantifying and incorporating observation uncertainties. The representation of the weighting factor ω in the cost function should be improved. In this study, only global performance measures were used to compare the flood maps. A possibility is to develop local performance measures that better capture both global agreement and localized mismatches between simulations and observations. Satellite-derived flood maps can be used to calibrate flood models when exploring the uncertainty space beyond floodplain roughness, including channel roughness, bathymetry, or inflow variability. Finally, alternative observation sources could be used, such as drone-based imagery or high-resolution optical satellites, which can offer better classification accuracy and reduced uncertainty.

Author contributions

JPT, SB, CG, VB, and FZ conceptualized the study. JPT developed the methodology, conducted the formal analysis and visualization, and prepared the original draft of the manuscript. SB, CG, VB, and FZ supervised the research and reviewed and edited the manuscript.

Code and data availability

The openTELEMAC software is freely available on a Gitlab server with a track of all the developments and a fixed branch for the v8p5 version used in the present study (https://gitlab.pam-retd.fr/otm/telemac-mascaret/-/tree/v8p5r0?ref_type=tags). The data assimilation module (ADAO) coupled with TELEMAC-2D is available as a Python library (<https://pypi.org/project/adao/>). For model construction, topographic data (for the Gironde and Lot-et-Garonne administrative départements) can be downloaded at <https://geoservices.ign.fr/rgealti>. The land use data from 2018 can be retrieved in a public catalog from the European Environment Agency. The hydrometric data at Tonneins, Marmande, and La Réole can be extracted from a public database (<https://www.hydro.eaufrance.fr/rechercher/entites-hydrometriques>) for stations O900001002, O909001001, and O919001001 respectively. The satellite observations used in the present study are available on ASF Data Search Vertex (<https://search.asf.alaska.edu/>) under scene name S1B_IW_GRDH_1SDV_20210202T175506_20210202T175531_025431_03077B_5B8F and S1A_IW_GRDH_1SDV_20210203T174749_20210203T174814_036429_0446C1_5E45.

Acknowledgments

This work was carried out as part of EDF R&D's MOISE-2 research project on river and coastal flood hazard assessment, whose support the authors gratefully acknowledge. This work was supported and funded by the French National Association of Research and Technology (ANRT) and EDF R&D with the Industrial Conventions for Training through REsearch (CIFRE grant agreement 2022/0972). The authors would like to gratefully acknowledge the open-source community, especially that of the openTELEMAC software (www.opentelemac.org) and the ADAO Python library (<https://pypi.org/project/adao/>).

A Additional results for the data assimilation on twin experiment

The results of the data assimilation for the four performance measures are reported in Figures 23 and 24 for subdomains 2 and 4, respectively.

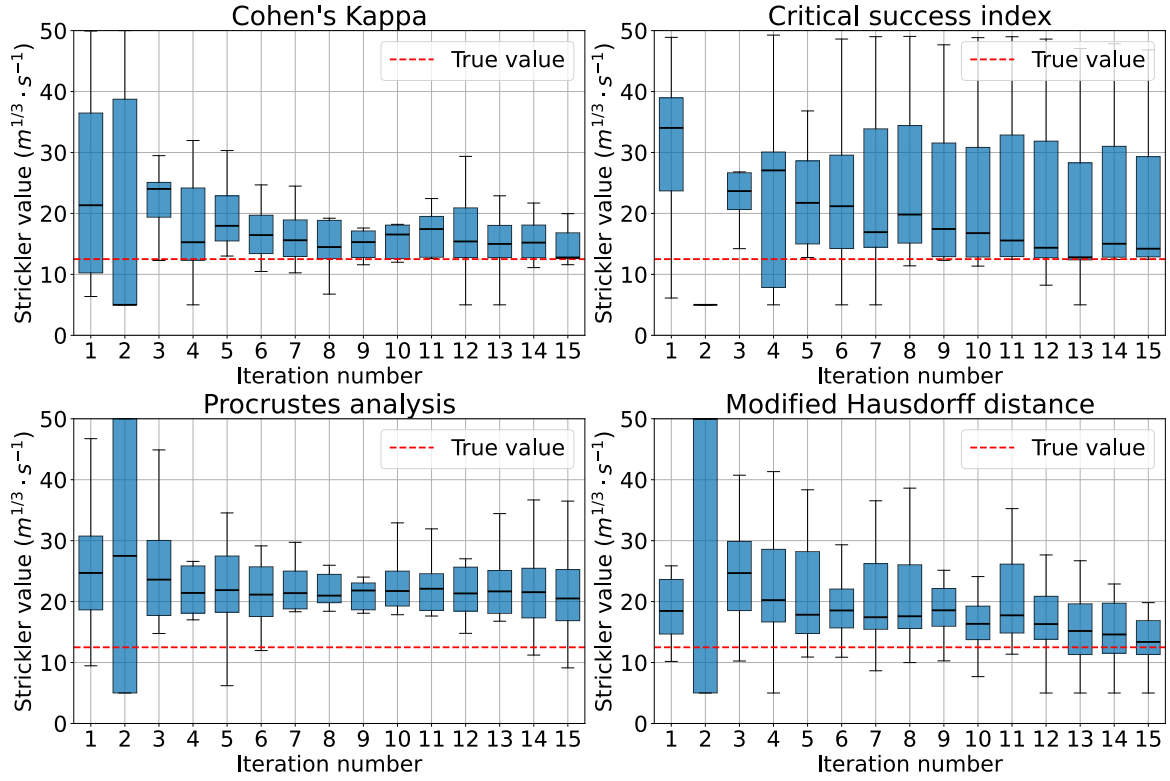


Figure 23: Estimates of the Strickler values for the twin experiment in subdomain 2 for four performance measures using 3D-Var. Boxplots show the variability of estimates across 10 initializations over iterations.

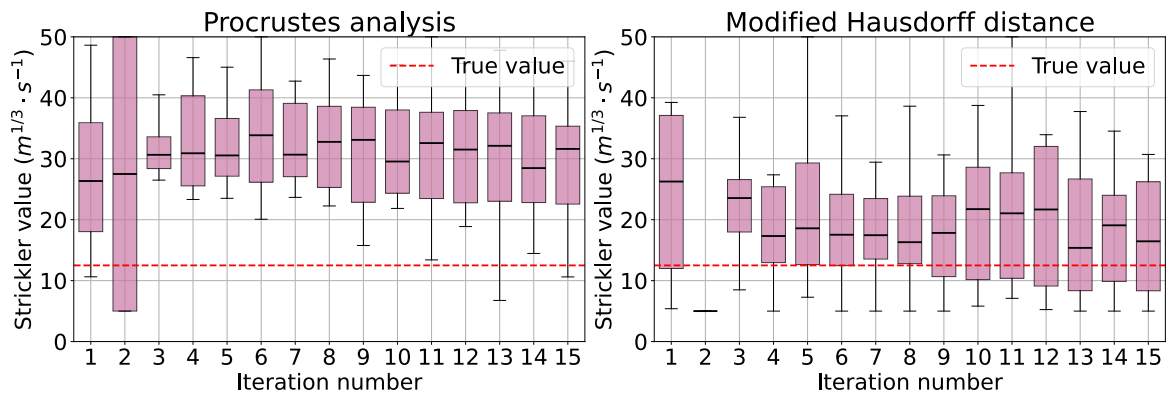


Figure 24: Estimates of the Strickler values for the twin experiment in subdomain 4 for two performance measures using 3D-Var. Boxplots show the variability of estimates across 10 initializations over iterations.

References

- Argaud, J.-P. (2019). User documentation, in the salome 9.3 platform, of the adao module for "data assimilation and optimization". Technical report, Technical report 6125-1106-2019-01935-EN, EDF/R&D.
- Aronica, G., Bates, P., and Horritt, M. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within glue. *Hydrological processes*, 16(10):2001–2016.
- Barnes, H. H. (1967). *Roughness characteristics of natural channels*. Number 1849. US Government Printing Office.
- Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A.-L. (2017). Openturns: An industrial software for uncertainty quantification in simulation. In *Handbook of uncertainty quantification*, pages 2001–2038. Springer.
- Besnard, A. and Goutal, N. (2011). Comparaison de modèles 1d à casiers et 2d pour la modélisation hydraulique d'une plaine d'inondation—cas de la garonne entre tonneins et la réole. *La Houille Blanche*, (3):42–47.
- Betterle, A. and Salamon, P. (2024). Water depth estimate and flood extent enhancement for satellite-based inundation maps. *Natural Hazards and Earth System Sciences*, 24(8):2817–2836.
- Bonafilia, D., Tellman, B., Anderson, T., and Issenberg, E. (2020). Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 210–211.
- Brakenridge, G., Tracy, B., and Knox, J. (1998). Orbital sar remote sensing of a river flood wave. *International journal of remote sensing*, 19(7):1439–1445.
- Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208.
- Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535.
- Chini, M., Hostache, R., Giustarini, L., and Matgen, P. (2017). A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):6975–6988.
- Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*. McGraw-Hill Book Company.
- Clement, M. A., Kilsby, C., and Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, 11(2):152–168.
- Cohen, S., Brakenridge, G. R., Kettner, A., Bates, B., Nelson, J., McDonald, R., Huang, Y.-F., Munasinghe, D., and Zhang, J. (2018). Estimating floodwater depths from flood inundation maps and topography. *JAWRA Journal of the American Water Resources Association*, 54(4):847–858.
- Cooper, E. S., Dance, S. L., García-Pintado, J., Nichols, N. K., and Smith, P. J. (2019). Observation operators for assimilation of satellite observations in fluvial inundation forecasting. *Hydrology and Earth System Sciences*, 23(6):2541–2559.
- Dalsasso, E., Denis, L., and Tupin, F. (2021). Sar2sar: A semi-supervised despeckling algorithm for sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4321–4329.
- Dasgupta, A., Hostache, R., Ramsankaran, R., Schumann, G. J.-P., Grimaldi, S., Pauwels, V. R., and Walker, J. P. (2021). A mutual information-based likelihood function for particle filter flood extent assimilation. *Water Resources Research*, 57(2):e2020WR027859.
- Dubuisson, M.-P. and Jain, A. K. (1994). A modified hausdorff distance for object matching. In *Proceedings of 12th international conference on pattern recognition*, volume 1, pages 566–568. IEEE.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, 32(2):407–499. With discussion, and a rejoinder by the authors.
- Goeury, C., Audouin, Y., and Zaoui, F. (2022). Interoperability and computational framework for simulating open channel hydraulics: Application to sensitivity analysis and calibration of gironde estuary model. *Environmental Modelling & Software*, 148:105243.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40:33–51.
- Grimaldi, S., Li, Y., Pauwels, V., and Walker, J. P. (2016). Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges. *Surveys in Geophysics*, 37(5):977–1034.
- Hervouet, J.-M. (2007). *Hydrodynamics of free surface flows: modelling with the finite element method*. John Wiley & Sons.

6.3 Complements: Twin experiment on a schematic test case

The influence of the choice of performance measures when assimilating flood maps in a 2D flood model with variational approaches was also evaluated on a schematic test case. The test case consists of a twin experiment on a meandering channel with a right and left floodplain. The aim is to calibrate the spatially distributed roughness, which is divided into three subdomains, as shown in Figure 6.1.

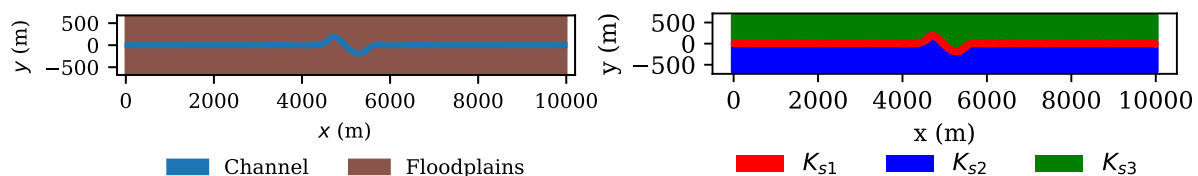


Figure 6.1: Illustration of the schematic test case for preliminary tests of the methodology presented in Section 6.2.

The methodology follows that described in Section 6.2, using a 3D-Var data assimilation algorithm to estimate the roughness coefficients by minimizing a cost function. In the cost function, the influence of the choice of performance measure and of the initial state θ^b on the estimation of the three parameters was evaluated. The simplified framework enabled the comparison of performance measures under idealized conditions.

The test case also highlighted the importance of the performance measure in the calibration process when using flood maps. It also demonstrated that the choice of initial parameter guesses can affect the outcome depending on the shape of the cost function.

The detailed results of these additional experiments are presented more thoroughly in Appendix A.1.

6.4 Summary

This chapter aimed to calibrate a spatially varying floodplain roughness parameter assimilating SAR-derived flood maps. We assimilated flood maps using a 3D-Var variational approach, where the aim is to minimize a cost function that combines background knowledge and an observation term, representing the misfit between model outputs and observations. We carried out assimilation on both a schematic and a real flood event scenario, highlighting the impact of the choice of performance measure on the estimation of the floodplain roughness parameter. In controlled twin experiments, the assimilation of flood maps enabled calibration of the flood model when using adapted performance measures, as described in Chapter 5. However, in a real case scenario, where observations are uncertain, the assimilation of flood maps for calibrating floodplain roughness was less successful due to large discrepancies between the simulated and observed flood maps, and because the uncertainty of the observed flood maps is hardly quantified for the tested performance measures. The degree of precision in current methods for generating flood maps from SAR imagery makes calibration of floodplain roughness difficult. However, this methodology can be used to calibrate the channel roughness, which is usually more influential on the flood maps. Then, observed flood maps, even with some uncertainties, can be useful for calibrating the model.

Chapter 7

Calibration of roughness field in 2D flood models with water depth fields assimilation

Résumé :

Ce chapitre vise à assimiler les champs de profondeur d'eau dérivés des observations satellitaires afin d'améliorer la calibration de la rugosité des plaines d'inondation dans les modèles hydrodynamiques 2D. Bien que l'estimation de la profondeur d'eau à partir d'observations satellitaires introduise plus d'incertitudes que les cartes d'inondation, elle peut simplifier le processus d'assimilation avec des observations de la profondeur d'eau plus proches des quantités simulées par les modèles hydrodynamiques.

Trois techniques d'assimilation de données (3D-Var, filtre de Kalman d'ensemble, filtre particulaire) sont testées sur une étude de cas schématique. L'influence des erreurs d'observation, de l'échantillonnage spatial et de la résolution temporelle sur la rugosité calibrée de la plaine d'inondation est évaluée.

La méthodologie est ensuite appliquée à un événement d'inondation réel sur la Garonne en France en utilisant l'algorithme 3D-Var. Tout d'abord, une expérience jumelle est utilisée pour évaluer comment les erreurs d'observation et la distribution spatiale des observations affectent les résultats. Enfin, la calibration de la rugosité de la plaine d'inondation est effectuée à l'aide d'observations de profondeur d'eau dérivées de cartes d'inondation obtenues par satellite. Les résultats indiquent que, malgré les incertitudes, la calibration de la rugosité de la plaine d'inondation peut être améliorée.

Abstract:

This chapter explores the assimilation of water depth field observations derived from satellite observations to improve the calibration of floodplain roughness in 2D hydrodynamic models. Although water depth estimation from satellite observations introduces more uncertainties than flood maps, it can simplify the assimilation process with water depth observations closer to the quantities simulated by hydrodynamic models.

Three data assimilation techniques (3D-Var, EnKF, and PF) are tested on a schematic case study. The influence of observation errors, spatial sampling, and temporal resolution on the calibrated floodplain roughness is evaluated.

The methodology is then applied to a real flood event on the Garonne River in France using the 3D-Var algorithm. First, a twin experiment is used to assess how observation errors and the spatial distribution of observations affect the results. Finally, floodplain roughness calibration is conducted using water depth observations derived from satellite-derived flood maps. The results indicate that, despite uncertainties in the simulated water depth, calibrating floodplain roughness can still be improved.

Contents

7.1	Motivations	206
7.2	Water depth assimilation on a schematic test case	206
7.3	Water depth assimilation on Garonne River	217
7.3.1	Water depth observations	217
7.3.2	Sensitivity analysis	219
7.3.3	Data assimilation	220
7.4	Summary	226

7.1 Motivations

In Chapter 6, we assimilated flood maps to infer a spatially distributed roughness field. When using flood maps derived from satellite observations, we observed the difficulty of calibrating the roughness parameter using global performance measures that compare simulated to observed flood maps. In Chapters 2 and 4, we showed that water depth fields can be extracted from satellite imagery by combining flood maps with a Digital Elevation Model.

Therefore, we evaluate the possibility of extracting water depths at various points within the flood extent and assimilate these point-wise observations on the floodplains. Although extracting water depth from satellite observations introduces additional uncertainty, the cost function is simpler to minimize, and water depth is a direct output of the simulation and a more natural target for assimilation. As in the previous chapter, this chapter aims to address the third research question introduced in Chapter 1:

Research Question 3

Is it possible to calibrate floodplain roughness in 2D hydrodynamic models using uncertain satellite observations?

Section 7.2 presents the assimilation of water depths at multiple locations across floodplains, under various noise levels, to calibrate a spatially variable roughness field in a schematic test case. The assimilation of water depths on this test case was conducted for three Data Assimilation techniques: 3D-Var, EnKF, and PF. This section is written in the form of a conference paper presented at the Telemac User Conference [Travert et al., 2024]. The same methodology is then applied to the Garonne River case study with 3D-Var only, using a twin experiment and water depths derived from satellite imagery to estimate the roughness coefficients.

7.2 Water depth assimilation on a schematic test case

Comparing Data Assimilation Algorithms For Parameter Estimation on TELEMAC-2D

Jean-Paul Traver^{1,2}, Cédric Goeury^{1,2}, Vito Bacchi¹, Fabrice Zaoui¹, Sébastien Boyaval^{2,3}

jean-paul.travert@edf.fr, Chatou, France

¹: National Laboratory for Hydraulics and Environment, EDF R&D, Chatou, France

²: LHSV, Ecole des Ponts, EDF R&D, Chatou, France

³: MATHERIALS, Inria, Paris, France

Abstract – Numerical models play an essential role in fluvial hydraulics for modelling flood dynamics. These models are often used as prediction tools to design effective mitigation measures or estimate possible damages due to overflowing for future events in specific areas. Before using these models, they need to be calibrated against observations to assess their accuracy for past and future events. In this study, we implement and compare in a realistic scenario different data assimilation algorithms for parameter estimation to evaluate which is most suitable for calibrating hydraulic models. In this work, the goal is to retrieve the friction, a distributed parameter in two-dimensional flood models. Water depth measurements at different places within the channel and floodplains are used as observations in Data Assimilation algorithms. The methodology is conducted on a twin experiment. Different levels of observation errors and various configurations of observations are tested, both temporally and spatially. The 3D-Var algorithm was more efficient for most configurations for identifying the parameters.

Keywords: Data Assimilation, Parameter Estimation, Roughness.

I. INTRODUCTION

During flood events, it is essential to predict water depths accurately. For floods, a common choice is to perform a numerical simulation with 2D Shallow Water Equations (SWE) models. However, uncertainties arise when constructing a numerical model with these equations for real cases. The primary sources of uncertainties in flood modelling are the model inputs (boundary conditions, initial conditions, topography, and bathymetric data) and the physical processes modelled by the friction [1].

To reduce uncertainties and accurately predict floods, models are calibrated. Calibration aims at adjusting uncertain physically based parameters (e.g., friction values) to fit measurements. For a long time, measurements were only water levels from stage gauging stations, gauge leases, or aerial photographs. Still, satellite imagery can now get more distributed data on the floodplains. These measurements can be water depths, flood contour, and velocities, but we will focus in this work on water depths. Using distributed water depths for calibrating models may lead to significant improvements compared to using water depths only in the channel.

The water depths and flood extent are usually very sensitive to the inlet discharge [2]. A stage gauging station at the inlet of the domain standardly allows one to reduce the uncertainty on the boundary and initial conditions strongly. Topography can now be known with a high degree of precision using modern numerical terrain models. Bathymetric data can be similarly known at the price of costly terrain measurements: we assume they are certain. So, the friction coefficients account for most of the remaining uncertainties which could be reduced by e.g., new data such as satellite images [2].

Data Assimilation (DA) aims to find the best compromise between observations and numerical models. This subfield of inverse problems, first applied to numerical weather prediction, is widely used in geosciences and hydraulics [3].

This paper uses DA for inverse parameters estimation from observations and the outputs of a numerical model (e.g., water depths).

This paper aims to compare and implement different algorithms of DA with parallelisation capabilities working with the TELEMAC-MASCARET software suite (www.opentelemac.org) and especially TELEMAC-2D. Most DA algorithms are designed for real-time updating of the states, but here we seek to evaluate parameters (distributed friction) fixed at the start of a simulation.

Section II mathematically sets the parameter estimation problem. Sections III, IV, and V present various DA algorithms to tackle the issue. Section VI presents numerical results to compare the various algorithms in a schematic test case for a flood event under different configurations. Finally, in Section VII, the conclusions and perspectives of this work are presented.

II. PROBLEM SETTING

Two significant components define the inverse problem. On the one hand, uncertain measurements (e.g., water depths), and on the other hand, the numerical model simulating hydraulic states. Our goal is to estimate the distributed friction parameters of the model through a comparison between water depth measurements and the outputs generated by the numerical model.

A. Setting of the parameter estimation problem

The evaluation of the model parameters uncertainties relies on limited information from the initial estimation of the parameters (including physical laws, previous simulations, expert opinions, etc) and imperfect measurements.

Let $\mathcal{M}$ be the non-linear numerical model depending on some control parameters θ . Let $\mathbf{x}_k \in \mathbb{R}^{N_x}$ be the state vector i.e. the hydrodynamic variables at a time k (the water depths and velocities computed in the numerical model). The state vector $\mathbf{x}_k$ satisfies (1):

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \theta) + \boldsymbol{\eta}_k, \quad (1)$$

where $\theta \in \mathbb{R}^{N_p}$ are the model parameters (distributed friction in our case), $\mathcal{M}_{k:k-1}: \mathbb{R}^{N_x \times N_p} \rightarrow \mathbb{R}^{N_x}$ is the model from time $k-1$ to k , and $\boldsymbol{\eta}_k \in \mathbb{R}^{N_x}$ are the model errors.

Our objective is to estimate the parameters vector θ knowing the observations $\mathbf{y}_k^{obs}$. We consider that these observations are noisy and are available at discrete times k . Considering that the errors are additive, the observation model is given in (2):

$$\mathbf{y}_k^{obs} = \mathcal{H}(\mathbf{x}_k) + \boldsymbol{\epsilon}_k, \quad (2)$$

where $\mathbf{y}_k^{obs} \in \mathbb{R}^{N_{obs}}$ are the observations, $\mathcal{H}: \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_{obs}}$ is an observation operator mapping (selection, linear, etc) from model to observational space and $\boldsymbol{\epsilon}_k \in \mathbb{R}^{N_{obs}}$ are the errors of observation. In most DA algorithms, we consider independent and identically distributed errors $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ and $\boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$. In this work, it is assumed that the model errors are null with $\mathbf{Q}_k = \mathbf{0}_{N_x \times N_x}$.

From these two equations, the estimation problem is defined. In the parameter estimation problem, the goal is to improve the characterisation of parameters θ . The problem can be written in a probabilistic framework from the random nature of uncertainties. Applying the Bayes' theorem, it comes in (3):

$$p(\theta | \mathbf{y}^{obs}) = \frac{p(\mathbf{y}^{obs} | \theta) p(\theta)}{p(\mathbf{y}^{obs})}, \quad (3)$$

Equation (3) is at the core of the parameters estimation problem as it enables to estimate the distributions of the unknowns θ based on the observations $\mathbf{y}^{obs}$. It relies on the prior belief on the parameters $p(\theta)$, the likelihood of the observations to occur for different values of the unknowns $p(\mathbf{y}^{obs} | \theta)$ and the probability of the observations $p(\mathbf{y}^{obs})$ independently from θ .

B. Shallow Water Equations with friction parameter

All the simulations throughout this paper are conducted using the hydrodynamic module TELEMAC-2D solving the 2D SWE [4-5]. This model represents the $\mathcal{M}$ model from paragraph A.

The conservative form of the equations write as follows in (4), where the system unknowns are the velocities $\mathbf{u} = (u, v)^T$ along the x and y Cartesian coordinates, and the water depths

$h = \eta - z_b$ with η the free surface elevation and z_b the bottom elevation,

$$\begin{cases} \frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{u}) = 0 \\ \frac{\partial (h\mathbf{u})}{\partial t} + \nabla \cdot (h\mathbf{u} \otimes \mathbf{u}) = -g\nabla\eta + \frac{h\mathbf{F}}{\rho} + \nabla \cdot (h\nu_e \nabla \mathbf{u}) - \frac{\boldsymbol{\tau}_b}{\rho} \end{cases} \quad (4)$$

where $\mathbf{F}$ are external forces (e.g., Coriolis, centrifugal, surface tension ...), ν_e is the effective viscosity resulting from kinematic, turbulent and dispersive viscosity, and $\boldsymbol{\tau}_b$ is the bottom shear stress. The water density ρ and the gravitational acceleration g are considered constant. No external forces and viscosity effects are considered in this work except the bottom shear stress $\boldsymbol{\tau}_b$.

In the Shallow Water Equations, the equations need closure for the bed shear stress component $\boldsymbol{\tau}_b$ that is not analytically described as a function of h and $\mathbf{u}$. Different formulas are available in the literature, but the bed shear stress is often denoted as in (5),

$$\boldsymbol{\tau}_b = \frac{1}{2} \rho C_f |\mathbf{u}| \mathbf{u}, \quad (5)$$

where C_f is a dimensionless friction coefficient. The friction coefficient is based on empirical or semi-empirical formulas [6]. In these equations, different parameters are defined depending on the soil surface, the land use and other physical properties.

The most widely used empirical formula is the Strickler formula [7] defined as in (6),

$$V = K_s R_h^{\frac{2}{3}} \sqrt{S_0}, \quad (6)$$

where K_s is the Strickler coefficient (defined from tables, calibration ...), R_h the hydraulic radius, V the mean velocity of the cross section and S_0 the bottom slope. There is an equivalent formula called the Manning formula mainly used by the Anglo-Saxon community with $K_s = \frac{1}{n}$ instead, n being the Manning coefficient.

This term in the equation induces energy dissipation due to the friction representing the water flow resistance caused by underlying surfaces. The energy losses directly impact the hydrodynamic variables. The friction parameter accounts for physical considerations such as surface roughness (irregularities of the bottom of the flow), channel irregularities or seasonal changes with growing vegetation, and phenomena such as turbulence or local errors in the topography. Thus, it remains difficult to estimate this parameter on a physical basis, and it should be adjusted with a calibration process by comparing water elevation and/or flood extent.

In TELEMAC-2D the friction term is considered in the momentum equation from the SWE and is implicitly treated [4]. The two components of the bottom friction calculated in x and y directions are given in (7),

$$\begin{cases} \tau_{bx} = \frac{\rho u}{2} C_f(h, K_s) \sqrt{u^2 + v^2} \\ \tau_{by} = \frac{\rho v}{2} C_f(h, K_s) \sqrt{u^2 + v^2} \end{cases}, \quad (7)$$

where $C_f = \frac{2g}{h^{1/3} K_s^2}$ for the Strickler formula.

In the parameter estimation problem, the unknown parameters θ that we aim to evaluate are the distributed frictions K_s from (6). The hydraulic states $\mathbf{x}_k$ at time k are represented in the equations by the water depths compared at the position of water depth observations. For evaluating, the parameters based on the measurements, different DA algorithms are presented in the next sections.

III. VARIATIONAL APPROACH: 3D-VAR

The first option for estimating the parameters from uncertain measurements and numerical models is to use a variational approach. The goal is to evaluate the unknowns by minimising a cost function quantifying the misfit between model predictions and measurements.

Most DA methods assume that errors are approximated with Gaussian assumption. For parameter estimation, the most used function for this minimisation problem is defined in (8),

$$\begin{cases} J(\theta) = J_b(\theta) + J_o(\theta) \\ J_b(\theta) = \frac{1}{2}(\theta - \theta^b)^T \mathbf{B}^{-1}(\theta - \theta^b) \\ J_o(\theta) = \frac{1}{2}(\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\theta)))^T \mathbf{R}^{-1}(\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\theta))) \end{cases}, \quad (8)$$

where θ^b is a background knowledge (a first guess on the set of unknown parameters), and $\mathbf{B}$, $\mathbf{R}$ are the background and observation error covariance matrices respectively. The proposed parameter estimation method is an extension for non-linear models of the computation of the Maximum A Posteriori (MAP) of (3) without model errors. In the linear and Gaussian assumptions case, Equation (3) can be derived as $p(\theta|\mathbf{y}^{obs}) \propto e^{-J(\theta)}$. Indeed, considering a Gaussian approximation, the prior distribution can be written as

$$p(\theta) = \frac{1}{\sqrt{(2\pi)^{N_p} \det(\mathbf{B})}} e^{(-\frac{1}{2}(\theta - \theta^b)^T \mathbf{B}^{-1}(\theta - \theta^b))}. \quad (9)$$

As for the likelihood, using Gaussian assumption with (1) and (2), it can be written as

$$p(\mathbf{y}^{obs}|\theta) = \frac{1}{\sqrt{(2\pi)^{N_{obs}} \det(\mathbf{R})}} e^{(-\frac{1}{2}(\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\theta)))^T \mathbf{R}^{-1}(\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\theta))))}. \quad (10)$$

Finally, by combining (3), (9) and (10), we get (11) where the cost function appears:

$$p(\theta|\mathbf{y}^{obs}) \propto e^{-(J_b(\theta) + J_o(\theta))}. \quad (11)$$

Equation (8) is the traditional data assimilation cost function called 3D-Var [8-9]. The 3D-Var algorithm is

programmed without dynamic evolution, considering only one assimilation window. The method is exact in the case of a linear model $\mathcal{M}$.

The cost function $J(\theta)$ from equation (8) should be minimised using an optimisation method. A gradient descent method is used which is available in the ADAO data assimilation library. From all the optimisation methods, we resort to the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. The algorithm is based on the application of the Newton method to the gradient of $J(\theta)$. The algorithm is purely iterative as it explores a direction with a set of parameters and then updates the parameters before launching a new simulation.

IV. FILTERING APPROACHES: ENSEMBLE-BASED APPROACHES

Kalman filters are sequential DA techniques, and the uncertainty in the state/parameters evolves from one observation to another, unlike the 3D-Var approach.

The Kalman Filter (KF) gives an exact solution of the Bayes problem in (3) for Gaussian errors and linear model assumptions ($\mathcal{M} \cong \mathbf{M}$ and $\mathcal{H} \cong \mathbf{H}$ with $\mathbf{M}$ and $\mathbf{H}$ being the tangent linear models). There are two different steps in the algorithm. First, the system's future states and uncertainties are computed based on the current states and dynamic model. Secondly, an analysis step is carried out to refine the state estimate by combining the forecast and the observation with a linear combination of the available information. The solution is given in (12) and (13). More details on the derivation of the solution can be found in [10]. In the parameter estimation problem, we only update the parameters, not the states. The f and a superscript respectively denote for forecast and analysis.

Forecasting step:

During the forecast step, the goal is to propagate the system state using the dynamics of the system and evaluate the error covariance matrix of the forecast in (12).

$$\begin{cases} \mathbf{x}_k^f = \mathbf{M}_{k:k-1} \mathbf{x}_{k-1}^a \\ \mathbf{P}_k^f = \mathbf{M}_{k:k-1} \mathbf{P}_{k-1}^a \mathbf{M}_{k:k-1}^T + \mathbf{Q}_k \end{cases}. \quad (12)$$

Analysis step:

For the analysis, the goal is to find $\mathbf{L}_k$ and $\mathbf{K}_k$ (the Kalman gain) as $\mathbf{x}_k^a = \mathbf{L}_k \mathbf{x}_k^f + \mathbf{K}_k \mathbf{y}_k^{obs}$. We describe hereafter in (13) the analysis step for the optimal linear combination between the prior and the observations. In this case, it is the exact solution of (3) with $\mathbf{x}_k^a$ having minimum variance and being unbiased.

$$\begin{cases} \theta_k^a = \theta_k^f + \mathbf{K}_k(\mathbf{y}_k^{obs} - \mathbf{H}_k \mathbf{x}_k^f) \\ \mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T)^{-1} \\ \mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \end{cases}. \quad (13)$$

However, the KF does not handle well non-linear models and is computationally expensive for covariance matrix estimations. The Ensemble Kalman Filter (EnKF) is a possible extension with a reduced computation cost. Instead of having a single estimate of states/parameters, the EnKF uses an ensemble of possible states (called members) capturing the uncertainty in the estimation. The states are forwarded in time with the model (forecast) and then updated with the perturbed observations (analysis). The implementation is based on [11] ([open source access](#)) and summarised in Figure 1.

Forecasting step:

Each member $i \in [1, N_e]$ is forecasted from t_k to t_{k+1} with Equation (1) without model error, and then the covariance matrix $\mathbf{P}_{k+1}$ and mean ensemble for the parameters are computed in (14).

$$\begin{cases} \mathbf{P}_{k+1}(\boldsymbol{\theta}_{k+1}) = \frac{1}{N_e-1} \sum_{i=1}^{N_e} (\boldsymbol{\theta}_{k+1}^i - \overline{\boldsymbol{\theta}_{k+1}})(\boldsymbol{\theta}_{k+1}^i - \overline{\boldsymbol{\theta}_{k+1}})^T \\ \overline{\boldsymbol{\theta}_{k+1}} = \frac{1}{N_e} \sum_{i=1}^{N_e} \boldsymbol{\theta}_{k+1}^i \end{cases} \quad (14)$$

Analysis step:

To be in a comparable set-up as 3D-Var, the updating step is carried out only at the end of an observation window. The observation window is set to the simulation time, so the parameters (distributed friction) are not updated during a run.

First the Kalman gain $\mathbf{K}$ is computed below,

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1} \mathbf{H}^T (\mathbf{H} \mathbf{P}_{k+1} \mathbf{H}^T + \mathbf{R})^{-1}. \quad (15)$$

Then the parameters are updated accordingly to:

$$\boldsymbol{\theta}_{k+1}^{i,a} = \boldsymbol{\theta}_{k+1}^i + \mathbf{K}_{k+1} (\mathbf{y}_{k+1}^{obs} + \boldsymbol{\epsilon}_k^i - \mathbf{H} \mathbf{x}_{k+1}^i). \quad (16)$$

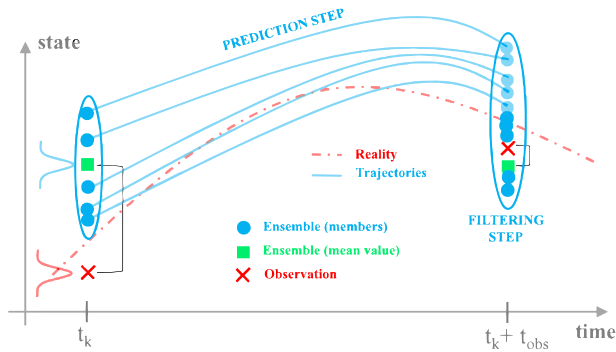


Figure 1. Sketch illustrating EnKF extracted from [11].

V. PARTICLE FILTER WITH RESAMPLING

Particle Filter (PF) solves Bayes' equation (3) based on a sequential Monte Carlo method without linear model or Gaussian error assumptions. The goal is to use a set of random samples (called particles) to get a direct solution of (3). The simplest version is the bootstrap Particle Filter introduced in [12]. From this basic version, we added resampling with

Kernel Density Estimation (KDE) for creating an iterative calculation chain.

The methodology is declined step by step. Each simulation can be parallelised, and all the particles can interact with each other at the observation time step.

1) Initialisation: Sampling procedure for the prior distribution

The first step is to sample $N_e \in \mathbb{R}$ particles $[\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^{N_e}]$. The parameter vector $(\boldsymbol{\theta}^i)_{1 \leq i \leq N_e} \in \mathbb{R}^{N_p}$ is a realisation of the prior distribution (the first guess) based on expert opinions, physical laws or previous simulations. The particles are drawn from $\mathcal{N}(\boldsymbol{\theta}^p, \mathbf{B})$. The probability density function of the parameter at time k is represented in (17):

$$p_k(\boldsymbol{\theta}) \cong \sum_{i=1}^{N_e} w_k^i \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k^i), \quad (17)$$

where $w_k^i \in \mathbb{R}^{N_p}$ is the weight of particle i (or likelihood of the particle) at time k and δ is a Dirac function ($\delta(0) = 1$). At initialisation, usually a uniform weight is considered with $\mathbf{w}_0 = [w_0^1, \dots, w_0^{N_e}] = [\frac{1}{N_e}, \dots, \frac{1}{N_e}]$.

2) Forecasting the particles

During the forecast step, each particle is forwarded with the numerical model until the next observation as in (18),

$$p_{k+1}(\boldsymbol{\theta}) \cong \sum_{i=1}^{N_e} w_k^i \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_{k+1}^i), \quad (18)$$

where $\boldsymbol{\theta}_{k+1}^i = \mathcal{M}_{k+1:k}(\mathbf{x}_k, \boldsymbol{\theta}_k) = \boldsymbol{\theta}_k^i$. In the case of the model parameters and friction coefficients, the parameters are not updated during a simulation. For the parallel implementation, each particle is forwarded until the next observation using *Runsize* processors. All the particles are run simultaneously when using $\text{Runsize} \times N_e$ processors otherwise the maximum of particles is launched and when the next time step of observation is reached the processors are used for other particles waiting in the queue, so that every particle reaches the observation step for the analysis step and before forwarding the model to the next observation.

3) Analysis step

During this step, the weights of the particles are updated according to (19),

$$w_{k+1}^{a,i} \cong w_{k+1}^{f,i} p(\mathbf{y}_{k+1}^{obs} | \mathbf{x}_{k+1}^i, \boldsymbol{\theta}_{k+1}^i), \quad (19)$$

where in this case $w_{k+1}^{f,i} = w_k^{a,i}$, because the forward model does not change the weights. The choice of the likelihood $p(\mathbf{y}_{k+1}^{obs} | \mathbf{x}_{k+1}^i, \boldsymbol{\theta}_{k+1}^i)$ is an open question. We choose a Gaussian likelihood (with assumption of independence between the observations, uncorrelated in time and in space) defined in (20), so that it can be compared with the other DA algorithms based on Gaussian errors:

$$p(\mathbf{y}_{k+1}^{obs} | \mathbf{x}_{k+1}^i, \boldsymbol{\theta}_{k+1}^i) = \prod_{l=1}^N \frac{1}{\sigma^l \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\mathbf{x}_{k+1}^{i,l} - \mathbf{y}_{k+1}^l}{\sigma^l} \right)^2}, \quad (20)$$

where N is the number of spatial observations, σ^l is the standard deviation of observation $l \in [1; N]$ considered constant for all spatial observations in the following.

Each particle weight is updated by comparing the water depths of the control vector of the numerical model and the observations. After that, all the processors communicate to normalise the weights so that $\sum_{i=1}^{N_e} w_k^{a,i} = 1$.

4) Resampling after each simulation

A well-known problem of PF is the curse of dimensionality, with one or few particles weights' largely dominating the others after some observations leading to poor distributions characterisation. At each new observation, and iteration, the parameters θ^i are resampled using KDE and the weights are reset to $1/N_e$. Then, after resampling we get back to step 2. The whole process can be visualised in Figure 2.

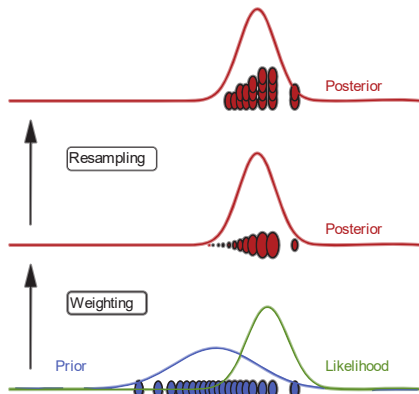


Figure 2. Schematic of PF with resampling extracted from [10].

VI. NUMERICAL STUDIES

In this section, the three DA algorithms are compared on a simple test case of flood under different configurations. The experiment is based on a numerically generated synthetic experiment, where the parameters' true values and the hydraulic variables' states are known. We resort to a synthetic experiment to have control over the test case and to focus on the numerical comparison and implementation of the algorithms.

A. Definition of the test case

1) Geometry

The test case is based on a 100 m wide and more than 10.25 km long channel with floodplains on both sides. On the floodplains, a constant slope along the y Cartesian coordinates is set to $S_{fp} = 0.002 \text{ m/m}$. Along the channel in the x direction, the slope is fixed to $S_{channel} = 0.001 \text{ m/m}$. The crest of the channel delineation with the floodplain has a slope $S_{b1} = 0.0012 \text{ m/m}$ until halfway of the reach and then $S_{b2} = 0.0008 \text{ m/m}$. Under these considerations, the channel depth is lower at $x = 5000 \text{ m}$, so it overflows in the middle without having much sensitivity to the boundary conditions. The channel section is trapezoidal, transitioning from the bank to the channel elevation on 10 m. Figure 3 describes the

geometric configuration at three cross-sections along the numerical domain.

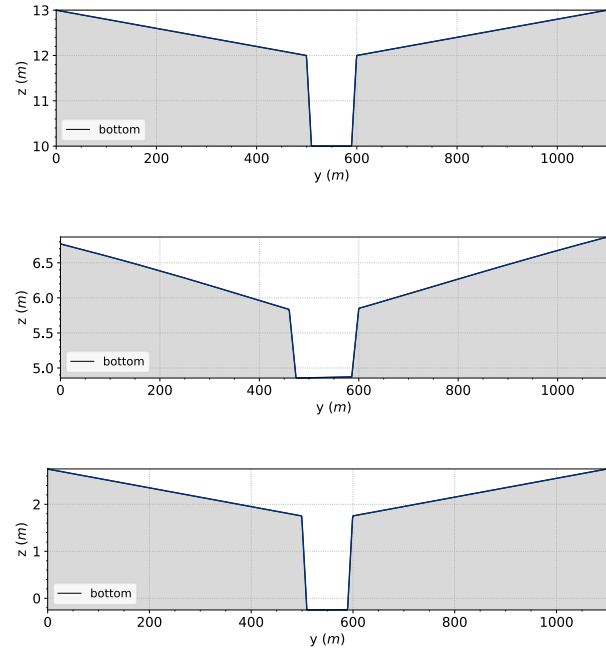


Figure 3. Elevation along a cross-section at $x = 0 \text{ m}$ (top), $x = 5000 \text{ m}$ (middle) and $x = 10000 \text{ m}$ (bottom).

The mesh is composed of 477 371 unstructured triangular elements. The number of elements may appear high for a test case, but the goal was to use a similar number of elements as in industrial application cases. The mesh is refined with different resolutions (prior mesh convergence has been carried out making sure the mesh is fine enough for our test case) from 2 m in the transition zone between the channel and the floodplains up to 15 m at the floodplain. As for the channel, the mesh resolution is of 5 m.

2) Initial conditions, inputs, and numerical considerations

For the upstream boundary condition of the model, we use a linear flow rate from $30 \text{ m}^3 \cdot \text{s}^{-1}$ to $620 \text{ m}^3 \cdot \text{s}^{-1}$ for the first five minutes. Then the discharge is constant at $620 \text{ m}^3 \cdot \text{s}^{-1}$. For the downstream condition, the depth is imposed with a linear rating curve with water depths between 0.5 m and 3.3 m for the respective discharges. The total simulation time is of 2 hours 30 (steady state is not reached). An overflow is modelled with changing water height during the simulation over the floodplain (by starting from a low discharge steady state).

3) Friction zoning and generated observations

In this case, three friction zones are considered with one friction for the channel and one friction zone on each side of the river channel. In the sloped regions between the channel and the floodplains, the friction is the same as the channel. The water depth observations are generated from the simulations with an initial guess of Strickler coefficients (in $\text{m}^{1/3} \cdot \text{s}^{-1}$)

equal to $[(K_s^b)_1; (K_s^b)_2; (K_s^b)_3] = [30; 20; 10]$ with the zoning defined in Figure 4.

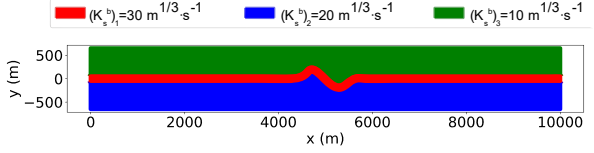


Figure 4. Roughness zones for the synthetic experiment.

From the reference simulation with the three friction zones, water depths are extracted at each mesh node at different times.

We consider two observation networks. A first approach OBS9 considers only 9 observation points (3 in the channel, 3 in the right floodplain, and 3 in the left floodplain). For the second configuration, two other points are added to the observation network (one in each floodplain) called OBS11 in the following. In Figure 5, the observation networks with 9 (OBS9) and 11 (OBS11) points are visualised.

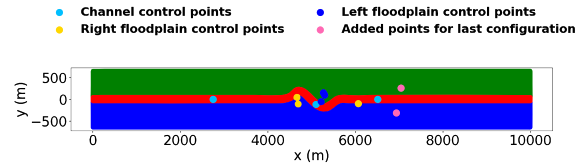


Figure 5. Observation network for the different DA experiments.

The water depths extracted from the synthetic experiment with the reference Strickler values are disturbed by considering different noise levels (adding a Gaussian noise to every observation). The water depths without noise at the different control points of OBS9 and OBS11 are visualised in Figure 6 for the reference simulation.

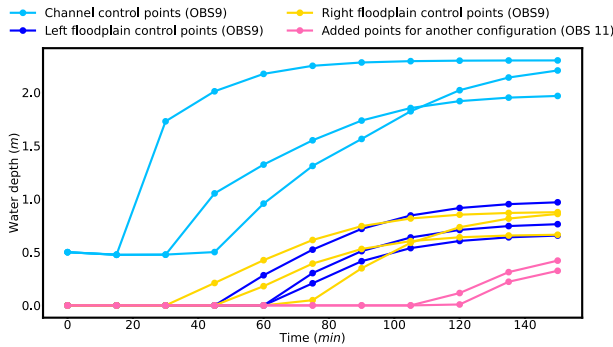


Figure 6. Water depths every 15 minutes at the OBS9 and OBS11 control points.

B. Software tools

Parameter identification under a probabilistic framework combined with optimisation and free-surface flows computation requires the call to different modules. Thus, external or internal libraries to TELEMAC-MASCARET are available to gain real-time interaction capabilities with the

TELEMAC-2D simulations, enabling on-the-fly adjustments and optimisations. In this work, the TELEMAC-2D solver (version V8P4) is coupled with TelApy, a module of the TELEMAC-MASCARET system, and ADAO, a Python data assimilation library (<https://pypi.org/project/adao/>).

1) The TelApy component of the TELEMAC system

One of the strengths of the TELEMAC-MASCARET software is the possibility to interact with the code during the simulations through Application Programming Interfaces (APIs) implemented for each module. The TelApy component is a Python source code wrapping the Fortran APIs [13-14]. The APIs provide the user with flexibility by allowing the control of the simulations while they are running. At any time during the simulation, the user can stop the simulation and retrieve or change the variables of the simulations through a list of variables declared as pointers. Thanks to that component, it is easier to implement a complete and efficient chain of calculations for parameter estimation under uncertainties. This component has been widely used to implement data assimilation strategies for parameter estimation or conjointly with the data assimilation library presented hereafter. In this work, TelApy is used for the three presented algorithms.

2) The Data Assimilation library (ADAO)

The variational approach described previously has been used through the Data Assimilation library called ADAO [8]. The ADAO module aims to simplify and provide different data assimilation and optimisation methodologies while being able to interact with other modules or codes in Python or SALOME (a scientific computing environment developed by EDF and CEA for industrial applications from mesh to post-treatment). It is used for various applications in the engineering field and provides a ready-to-use interface for users with an adapted implementation for industrial cases.

C. Results

1) DA variational algorithm: 3D-Var

The initial state of the friction parameters is set to $\theta^b = [(K_s^b)_1; (K_s^b)_2; (K_s^b)_3] = [35; 12; 18]$ with $B = 100 \cdot I_3$ as background covariance errors. The parameters are constrained between $5 \text{ m}^{1/3} \cdot \text{s}^{-1}$ and $90 \text{ m}^{1/3} \cdot \text{s}^{-1}$ and the differential increment is imposed arbitrarily to 10^{-2} for computing the observation operator approximated derivatives.

The results are not sensitive to the initial state, and the approach on the test case was robust for any initial state with a similar number of iterations for convergence.

The 3D-Var algorithm is studied with different observation configurations (OBS9 and OBS11), different noising of the observations ($\sigma^{obs} = 0.05 / 0.1 / 0.2 \text{ m}$) and different numbers of observation times (2, 10).

The roughness at the floodplains has low sensitivity to the water depth observations. Therefore, the water depths are noised with a small standard deviation of $\sigma^{obs} = 0.05 \text{ m}$ with $R = (\sigma^{obs})^2 \cdot I_{N_{obs}}$. Two observation times ($\#T_{obs} = 2$) are considered with $T = [1h15, 2h30]$. We compare configuration

OBS9 (9 points) and configuration OBS11 (11 points) in Figure 7.

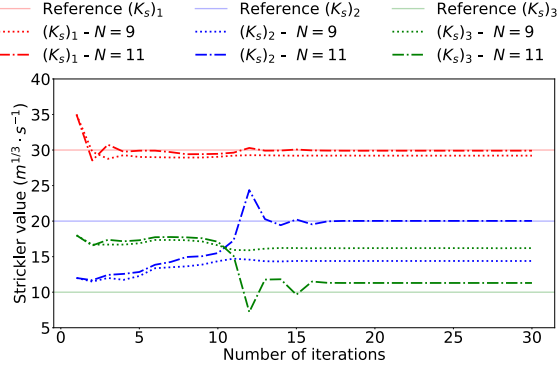


Figure 7. Comparison of parameter estimation for the three friction zones with two different observation networks OBS9 and OBS11 and $\sigma^{obs} = 0.05 m$.

In Figure 7, after a few iterations the channel roughness value is identified. For the floodplain coefficients it needs more iteration because of the lower sensitivity on the water depths. For OBS9, the floodplain parameters could not be identified correctly. Adding two more observation points in the floodplains improved the identification especially for $(K_s)_2$. It is essential to choose the observation points carefully to identify the parameters of the floodplains. $(K_s)_3$ is more rough leading to lower velocities and slower changes. This coefficient is more difficult to identify because of lower sensitivity (for information, switching the roughness values for $(K_s)_2$ and $(K_s)_3$ leads to similar results with a better identification of the rougher zone). For further investigations OBS11 is used.

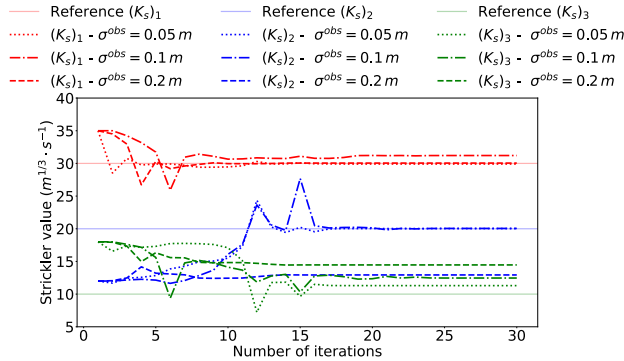


Figure 8. Comparison of parameter estimation for the three friction zones with OBS11 with $\sigma^{obs} = 0.05 m$, $\sigma^{obs} = 0.1 m$ and $\sigma^{obs} = 0.2 m$.

In Figure 8 parameter estimation is compared for different noising values of the observed water depths. Up to $\sigma^{obs} = 0.1 m$ the parameters are identified, but the floodplain parameters are far from the reference for higher values. With higher standard deviation on the measurement errors, we get almost no information from the measurement as a change in the roughness leads to low water changes compared to measurement errors.

Another test was to add times of observations. Instead of $T = [1h15, 2h30]$, observed water depths are acquired every 15 minutes (10 observations). In Figure 9, the identification of the parameters is slightly comparable by adding more temporal observations. The channel is closer to the reference as well as $(K_s)_3$. For $(K_s)_2$ the identification is a bit worse. For that test case, adding temporal observations does not change much the identification of the parameters. Considering the Gaussian uncertainties on the measurements, the differences between the two experiments are insignificant.

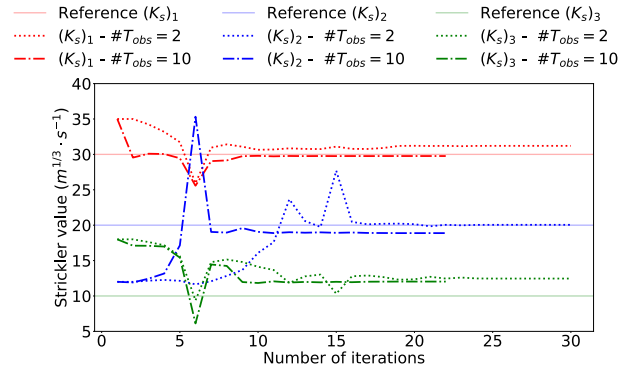


Figure 9. Comparison of parameter estimation for the three zones with configuration 2, $\sigma^{obs} = 0.1 m$ and different number of temporal observations.

With only two temporal observations, it is possible to evaluate the roughness values because the water depths at these time steps are sensitive to all the roughness zones. The floodplain values are not identified for other sets of two temporal observations as $T = [1h00, 2h00]$ instead of $T = [1h15, 2h30]$. It is probably because at these time steps the water depths for the added points for OBS 11 configuration are zero (no sensitivity yet from the roughness), as shown in Figure 6.

In every case, the cost function no longer reduces after 20 iterations (one iteration corresponds to four simulations for exploring the three parameters). When adding temporal observations, the convergence is faster with more information available. For this case, twenty iterations represent 2h40 of computing time on 96 processors (Intel(R) Xeon(R) Platinum 8260 CPU @ 2.40GHz) on the CRONOS Cluster from EDF.

2) Ensemble Kalman Filter (EnKF)

For EnKF the error is propagated on the parameters (distributed friction), and they are updated at the end of an assimilating window (smoothing processing). The assimilation window is set to the simulation time to be comparable with 3D-Var. According to the 3D-Var investigations the OBS11 observation network with a noise value of $\sigma^{obs} \leq 10 cm$ for the observations are chosen for the comparison. EnKF relies on the number of members chosen to represent the possible states and the number of temporal observations. The initial state members are drawn from a multivariate Gaussian $\mathcal{N}(\theta^b, \mathbf{B})$ and the parameters are bounded as in 3D-Var. We explore the role of the number of temporal observations and the ensemble size.

The algorithm is inefficient at evaluating the parameters for only two temporal observations. This can be explained by the fact that it is a sequential DA algorithm. With only two observations, there are not enough steps to converge towards the reference parameters. Therefore, 10 observations every 15 minutes are used instead. Figure 10 describes the evolution of the mean Strickler values for 25 and 100 ensemble sizes and configuration OBS11 with observation noise of $\sigma^l = 0.1 m$ and 10 temporal observations.

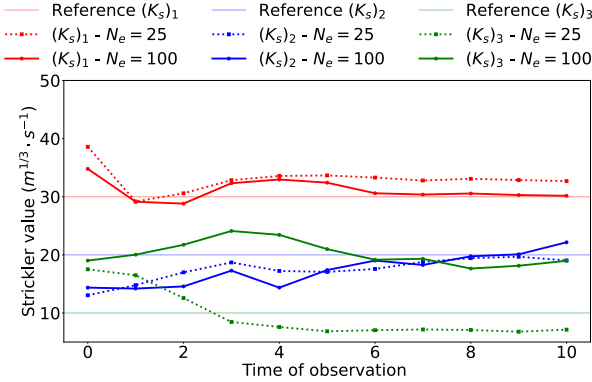


Figure 10. Evolution of the mean parameters' values for the three zones with configuration OBS11, $\sigma^{obs} = 0.1 m$ for different ensemble sizes (25 and 100) of EnKF.

For EnKF with several temporal observations, it is possible to identify the coefficients on the floodplains. As for 3D-Var the channel and the right floodplain are well identified at the end of the simulations for $N_e = 100$. For a smaller ensemble size N_e , the identification of the channel value is not as good, and it impacts the floodplain identification. Indeed, as the Strickler value is higher for the channel, the floodplain parameter values are lower to compensate. Having enough members is essential to sample correctly the search space. If not, the algorithm may diverge to another solution that is considered equally good, considering the uncertainties.

The algorithm works better for steady state as described in [11]. However, by doing an analysis step only at the end of the assimilation window, it can work for non-steady-state simulations if the parameters are not updated during the simulation.

Attempts to add more temporal observations did not lead to a better identification of the parameters. As for the comparison with 3D-Var, the identified coefficients for $(K_s)_1$ and $(K_s)_2$ are similar and the same identification problem for the third zone is encountered. The disadvantage of EnKF compared to 3D-Var is the need for more observations to get similar results, at least in this test case.

3) Particle Filter

Again, the OBS11 observation network with a noise value of $\sigma^{obs} = 10 cm$ for the observations and 2 observation times are chosen. The particles are drawn from a multivariate Gaussian $\mathcal{N}(\theta^b, B)$ and the parameters are bounded as in 3D-Var.

Without resampling we identify only the channel roughness as the water depth is very sensitive to the chosen control points to channel roughness and not floodplains (see Figure 11). The weight of a particle is unique, so it privileges the identification of the channel parameter. Furthermore, in most cases, the weight of one particle dominates the other, so the posterior distribution is not very well described.

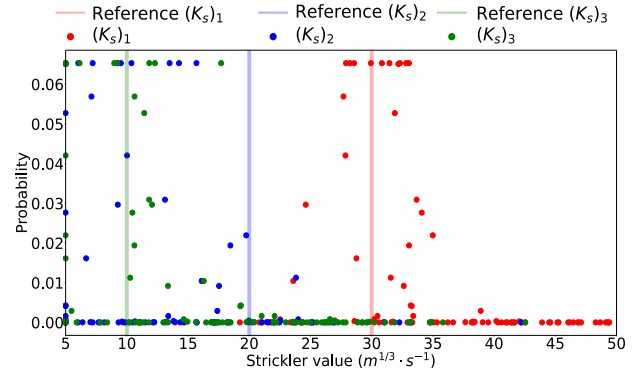


Figure 11. After first iteration of the Particle Filter with 100 particles.

Thus, we apply the Particle Filter iteratively by using the new posterior distribution for each parameter and use it as prior for a new iteration after resampling. After the first iteration, the variance of the Strickler parameters is low compared to the floodplains, so it is better identified. Then, at the next iteration, with resampling, the search space is drastically reduced for the channel parameter value compared to floodplains. The identification of the floodplains values is easier. Therefore, for the channel, which was greatly identified, the variance is low, and in the next iterations, the prior distribution is more centered on the predicted channel value, and thus it can focus on the identification of the coefficients from the floodplains by having high-density sampling around the channel reference value with low variance. In Figure 12 the channel parameter is well identified with a distribution around it (underestimating slightly the parameter), and the floodplains posterior distributions for the roughness are not centered on the reference but are near the value of the reference.

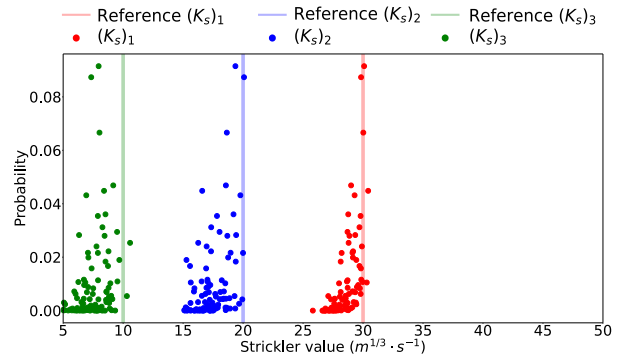


Figure 12. After five iterations of the Particle Filter with 100 particles.

The number of particles at the initialization also plays a role in exploring the set of parameters. When using more particles at the start, more sets of parameters are explored, so that the resampling might be more efficient.

However, with the iterative process and resampling, if the set of samples grasps the suitable parameters at first, it can converge quickly and give similar results. In Figure 13, it can be seen with the same number of iterations and only 25 particles instead of 100 showing the importance of the initial sampling. For other initial samples, 25 particles may not be enough to converge to the reference values.

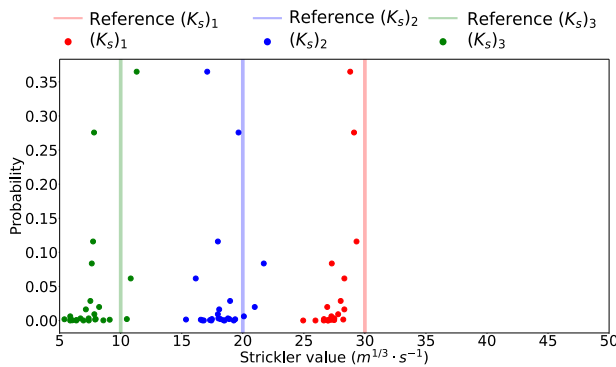


Figure 13. After five iterations of the Particle Filter with 25 particles.

When having more temporal measurements, floodplain values are difficult to find in this inverse problem because for most observation time steps, only the channel roughness influences the water depths at the control points. Furthermore, looking more closely at the water depths, the tricky point is that different parametrisations of the roughness may lead to the same output when comparing observations and models; it is called equifinality [15].

VII. CONCLUSION

In this paper, different methodologies for parameter estimation were carried out on a simplified but real-world dimensioned test case. It shows the potential of using observations combined with numerical models to improve the characterisation of friction parameters for flood studies. The TELEMAC-MASCARET software linked with ADAO and TelApy can lead to a faster learning curve for the user for Data Assimilation. It is possible to implement other DA algorithms, as it was the case with the implementation of the Particle Filter with resampling and parallelisation capabilities.

The 3D-Var algorithm was more adapted for parameter estimation under uncertainties on the test case. It was shown that identifying the three parameters was possible up to a specific error of observations. When the number of temporal measurements increases, the performance decreases, so it might be interesting to compare it with a 4DVAR approach to consider the dynamic evolution of the model. Another finding is the importance of the observation network. Indeed, adding two observation points on the floodplains where the water height was more sensitive to the friction of the floodplains at specific time steps led to a most straightforward estimation.

As for the EnKF, it proves to be adapted when enough temporal observations are available. It should be used with a smoothing procedure by updating the parameters at the end of an assimilation window and not at each observation (to compare it with the other algorithms).

The Particle Filter reaches better results than the prior distribution and enables the practitioner to know the uncertainties through a posterior distribution of the parameters given the uncertainties. However, using a Particle Filter with resampling in an iterative process requires much computation. The resampling procedure must be improved, so that the algorithm keeps exploring a wider set of samples, as it tends to explore a very restricted range of values when converging. The hyperparameter in the resampling method may be adapted to reach better performance.

For the number of calls to the TELEMAC-2D code, between the Particle Filter, EnKF, and 3D-Var, the latter is more efficient in general. For similar results, 3D-Var needed around 80 calls to the code, while PF needed 500 calls for five iterations with 100 particles and 125 with 25 particles (5 iterations of the PF). It is difficult to compare the number of calls to the code between the two methods as it depends on the number of chosen particles for the PF and the initial sampling. For the EnKF around 100 calls to the code were necessary to identify the parameters, but it was working only with ten temporal observations instead of two for the 3D-Var. The advantage of the PF and EnKF is that we can launch the particles in parallel.

In perspective, the methodologies should be conducted on real-world applications to evaluate different DA algorithms with a complete review of the available parameters to evaluate their performance. In the PF there is a need to work on the resampling procedure and the definition of the likelihood between the observations and the model predictions.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the French National Association of Research and Technology (ANRT) for funding and the open-source community, especially that of the TELEMAC-MASCARET (hydrodynamics) and ADAO (data assimilation).

REFERENCES

- [1] P. Bates, D. Pappenberger, and R. J. Romanowicz, "Uncertainty in flood inundation modelling", *Applied uncertainty analysis for flood risk management*, 2014, pp. 232-269.
- [2] F. Pappenberger, K. J. Beven, M. Ratto and P. Martgen, "Multi-method global sensitivity analysis of flood inundation models." *Advances in water resources*, 2008, pp. 1-14.
- [3] M. Asch, M. Bocquet, and M. Nodet, "Data assimilation: methods, algorithms, and applications", *Society for Industrial and Applied Mathematics*, 2016.
- [4] J.-M. Hervouet, "Hydrodynamics of Free Surface Flows: Modeling with the Finite Element Method", *John Wiley & Sons, Ltd: Hoboken, NJ, USA*, 2007.
- [5] A. B. de Saint Venant, « Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit », *Rapport technique, Académie des sciences*, 1871.
- [6] H. Morvan, D. Knight, N. Wright, X. Tang and A. Crossley. "The concept of roughness in fluvial hydraulics and its formulation in 1d, 2d

- and 3d numerical simulation models”, *Journal of Hydraulic Research*, 2008, vol. 46, pp. 191-208.
- [7] R. Manning, J. P. Griffith, T. Pigot, and L. F. Vernon-Harcourt, “On the flow of water in open channels and pipes”, 1890.
 - [8] J.-P. Argaud, “User Documentation, in the SALOME 7.5 platform, of the ADAO module for Data Assimilation and Optimization”, EDF R&D report, 2016.
 - [9] O. Talagrand, “Assimilation of observations, an introduction (gtspecial issuelldata assimilation in meteorology and oceanography: Theory and practice)”, *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B), 1997, pp.191-209.
 - [10] A. Carrassi, M. Bocquet, L. Bertino, G. Evensen, “Data assimilation in the geosciences: An overview of methods, issues, and perspectives”, *Wiley Interdisciplinary Reviews: Climate Change*, 2018.
 - [11] F. Zaoui, C. Goeury, Y. Audouin, “Ensemble Integrations of Telemac-Mascaret for the optimal model calibration.” In XXVth Telemac & Mascaret User Club, 2018.
 - [12] N. Gordon, D. J. Gordon, A. F. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation”, *IEE proceedings F (radar and signal processing)*, vol.140, No.2, 1993, pp. 107-113.
 - [13] C. Goeury, Y. Audouin, F. Zaoui, “User documentation v7p3 of TelApy module”, 2017
 - [14] C. Goeury, Y. Audouin, and F. Zaoui, “Interoperability and computational framework for simulating open channel hydraulics: Application to sensitivity analysis and calibration” of Gironde estuary model. *Environmental Modelling & Software*, 2022.
 - [15] K. Beven, “A manifesto for the equifinality thesis”, 2006, pp. 18-36.

7.3 Water depth assimilation on Garonne River

The objective of this section is to assimilate water depth fields from a real case study on the Garonne River to improve the characterization of a spatially varying roughness field. The same study area and satellite observations as in Chapter 6 are used for the 2021 flood event. This aims to constrain the hydraulic model more by assimilating water depth fields derived from satellite observations, rather than flood maps. In particular, the spatial distribution of water depth measurements and the observation errors impact on the calibration were evaluated, since they had a significant impact on the Strickler estimates in the schematic test case in Section 7.2.

7.3.1 Water depth observations

As discussed in Section 7.2, the number and spatial distribution of water depth measurements play an important role in recovering the true parameter field with automatic calibration strategies. Then, in this study, two observation configurations are considered for the assimilation of water depth.

In the first configuration, a total of 100 observation locations were selected randomly within the segmented flooded areas derived from the two satellite acquisitions. For the two observations, the locations of the randomly selected points are reported in Figure 7.1.

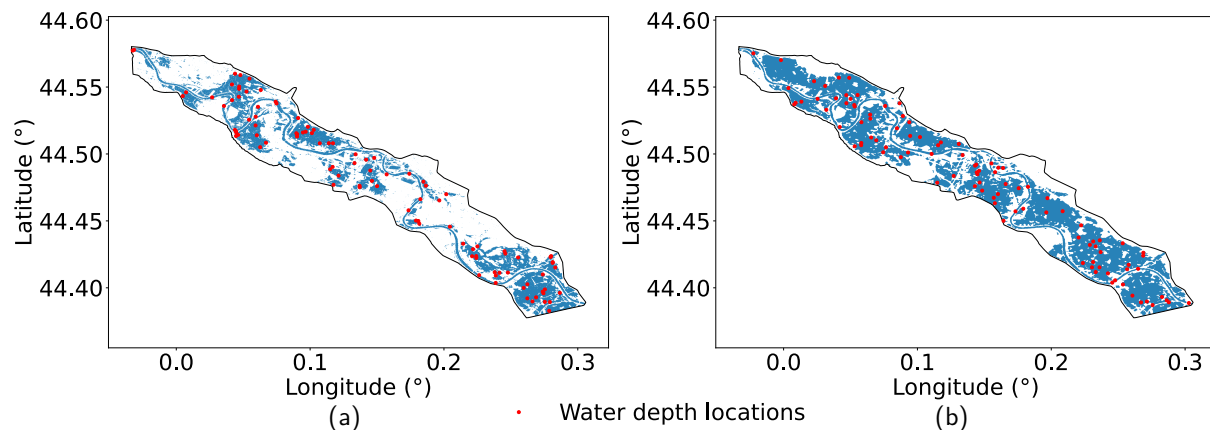


Figure 7.1: Selected water depth observation locations for configuration 1 on the Garonne River: (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

In the second configuration, a reduced number of 20 observation locations were considered and are shown in Figure 7.2. In contrast to the first configuration, these locations were selected manually based on hydraulic considerations. The evaluation of water depth estimation methods using flood maps in Chapter 4 revealed that larger errors tend to occur in regions with steep topography or complex irregular flood boundaries, for all water depth estimation methods (FwDET, FLEXTN, and the cross-section approach). The 20 locations from configuration 2 were placed in flatter areas where the flood extent appears relatively smooth. Thus, this configuration focuses on regions where water depth estimation is more accurate with reduced observation errors. The impact of these configurations on the estimation of the roughness field was evaluated.

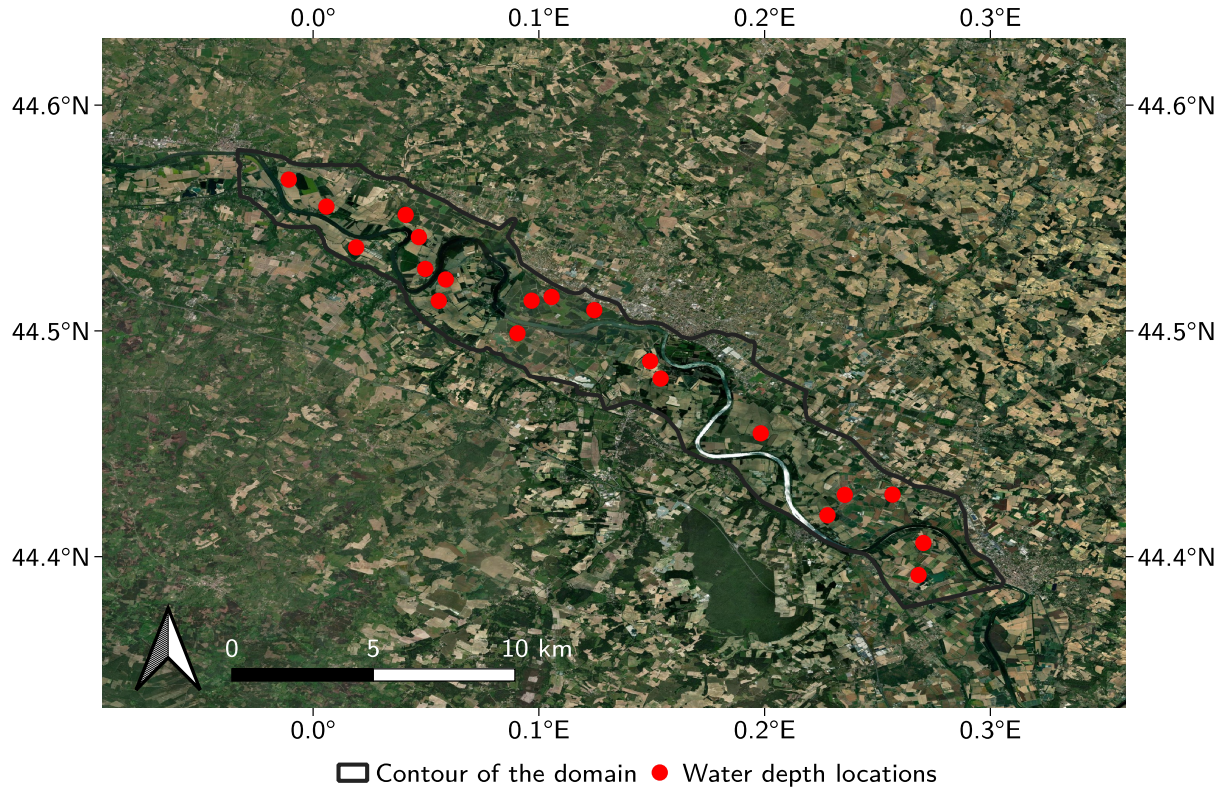


Figure 7.2: Selected water depth locations for the second configuration for water depth assimilation.

Similar to the assimilation of flood maps on the Garonne River in Chapter 6, two approaches were used for generating the observations for the assimilation. First, for the two configurations described in this section, the water depth fields were generated using the simulated results from a twin experiment. The second approach used water depth fields generated based on SAR observations of the flood event. The processing of the observation for both cases is described in Sections 7.3.1.1 and 7.3.1.2.

7.3.1.1 Synthetic Observations from a Twin Experiment

The hydrodynamic model is simulated for a fixed reference Strickler value, using the same reference values as the twin experiment for the assimilation of flood maps in Chapter 6. The simulated water depths at the two observation times (February 2, 2021, at 17:56 UTC, and February 3, 2021, at 17:48 UTC) are retrieved. These water depth fields are interpolated at the 100 observation locations for configuration 1 and 20 observation locations for configuration 2.

7.3.1.2 Water Depth Observations derived from Satellite Imagery

The water depths at the observation locations were generated based on SAR imagery. For each satellite observation, a flood map was generated, consisting of a grid of pixels of size $N_x \times N_y$. In this chapter, flood maps were generated using local thresholding techniques, as described in Section 6.2. The flood maps are combined with a DEM to estimate a water depth field for the flooded regions. Based on the analysis of the water depth estimation methods applied to the Garonne River case, we used the FLEXTM method [Betterle and Salamon, 2024] with the optimized hyperparameter configuration presented in Chapter 4. More details on the FLEXTM method can be found in Section 4.2. In Figure 7.3, the

generated water depth fields are reported for the two acquisition times during the 2021 flood event. The water depths for the configuration with 20 or 100 water depth sampling locations were then extracted from these water depth fields.

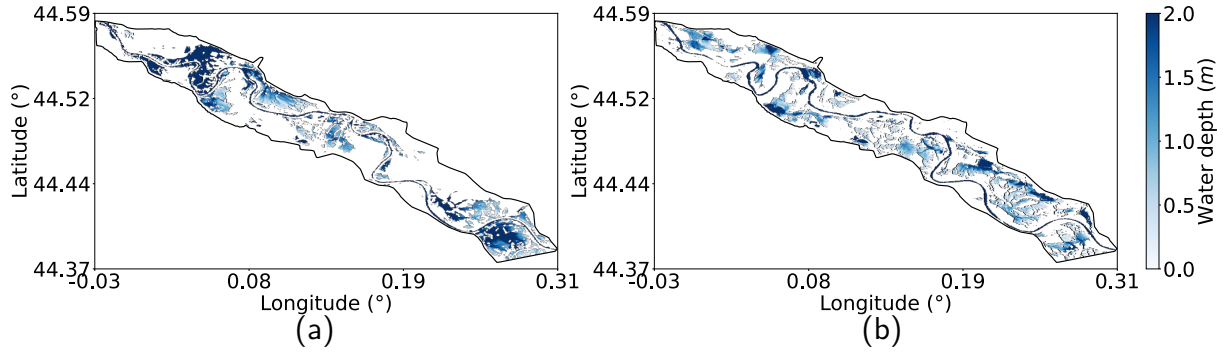


Figure 7.3: Generated water depth fields with FLEXTH method on the Garonne River: (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

7.3.2 Sensitivity analysis

The number of parameters was reduced using sensitivity analysis techniques described in Chapter 3. At the two observation times, a sensitivity study was conducted on the simulated water depth at the observation locations defined in Section 7.3.1. Total-order Sobol indices (TSI) are computed for each scalar output (i.e., each water depth location and observation time) with PCE as described in Sections 3.2.3.2 and 6.2. The surrogate models to compute the Sobol indices were trained on 6,000 simulations and validated on an independent set of 2,000 simulations.

Figure 7.4 presents the total-order Sobol indices computed across the floodplain subdomains for a single observation location in configuration 1, at the two observation times.

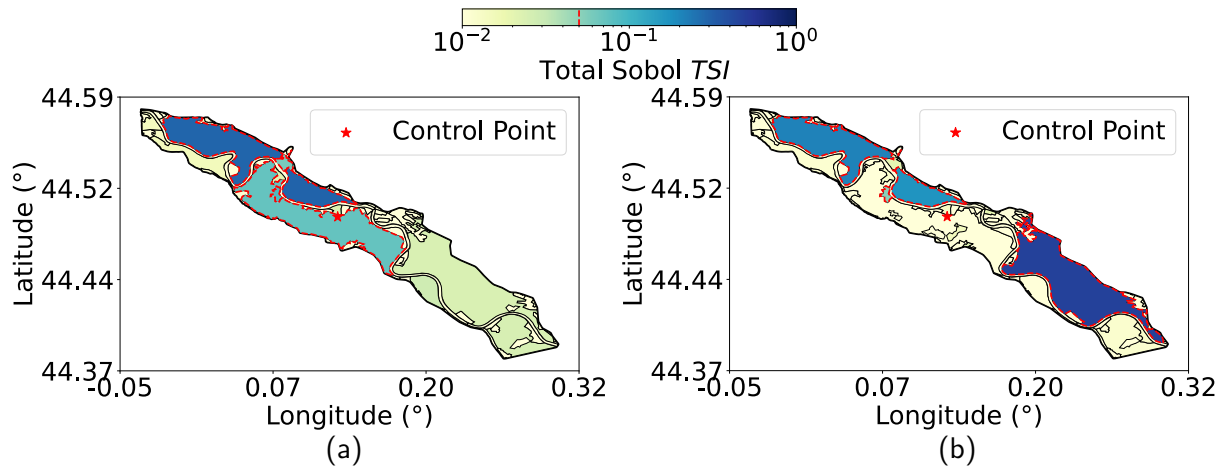


Figure 7.4: Total-order Sobol indices computed on the 92 floodplain subdomains for the water depth simulated at one location for (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

At the first observation time, the water depth at this point is mainly influenced by the Strickler value of the downstream subdomains. For the second observation time, again, three subdomains have high Sobol with two influential subdomains downstream and one upstream. The influential subdomains are mainly the larger ones in the study area,

which have a greater impact on the propagation of the flood. The spatial distribution of sensitivity highlights how the influence of roughness parameters can vary significantly over time, likely due to the dynamics of the flood and the direction in which it is propagating.

In the assimilation process, all the observed water depth measurements are assimilated. Then, the influential parameters for at least one of the water depth locations are conserved for the data assimilation step. Figure 7.5 summarizes the subdomains with $TSI > 0.05$ for at least one of the locations for the two acquisition times for configuration 1. For February 2, 2021, five subdomains are influential, and four are influential on February 3, 2021. For the DA step described in Section 7.3.3, the five subdomains are calibrated since they are either influential on February 2 or February 3 water depth outputs. Therefore, the number of parameters to calibrate is reduced in the DA step. For the second configuration with 20 observation locations, the same five subdomains are influential and calibrated.

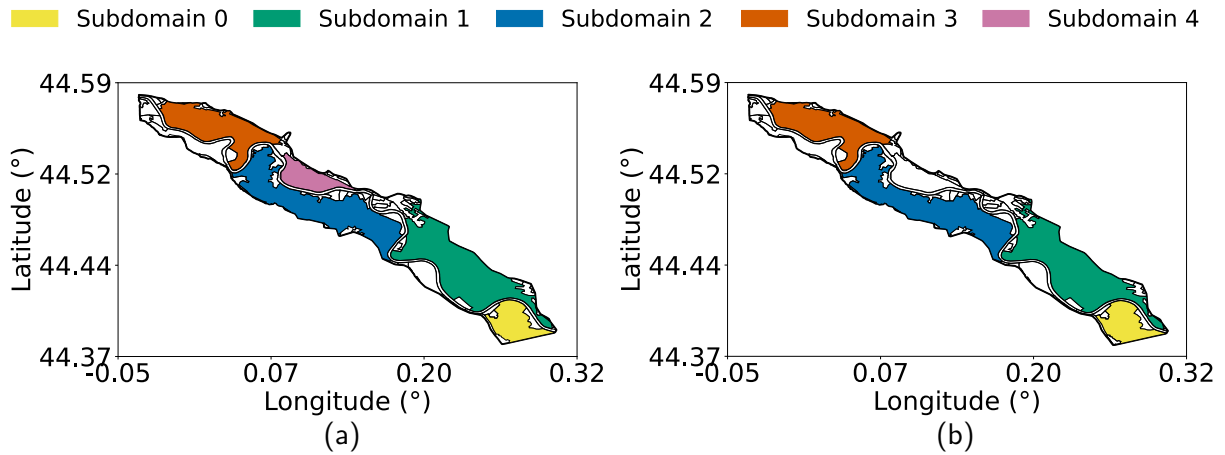


Figure 7.5: Subdomains with total-order Sobol indices higher than 0.05 for (a) February 2, 2021, at 17:56 UTC, and (b) February 3, 2021, at 17:48 UTC.

7.3.3 Data assimilation

The assimilation of the water depth field is carried out using the classical 3D-Var described in Section 3.6. We recall the cost function to minimize writes as:

$$\begin{aligned} J(\boldsymbol{\theta}) &= J_b(\boldsymbol{\theta}) + J_o(\boldsymbol{\theta}) \\ &= (\boldsymbol{\theta} - \boldsymbol{\theta}^b)^T \mathbf{B}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^b) + (\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\boldsymbol{\theta})))^T \mathbf{R}^{-1} (\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\boldsymbol{\theta}))), \end{aligned} \quad (7.1)$$

where $\boldsymbol{\theta}$ is the roughness parameter vector to calibrate (reduced set from sensitivity analysis), $\boldsymbol{\theta}^b$ is the background knowledge, $\mathbf{y}^{obs}$ is the vector of water depth observations at either 20 or 100 locations for the two acquisition times and $\mathcal{H}(\mathcal{M}(\boldsymbol{\theta}))$ is the projection of the TELEMAC-2D outputs to the observation space to compare the simulated water depths with $\mathbf{y}^{obs}$. The 3D-Var was initially designed for static problems; however, we consider here that all observations are assimilated as if they were observed at the initial time $t = 0$.

The cost function is minimized using the L-BFGS-B algorithm [Byrd et al., 1995] and via finite differences to compute the gradient. The number of maximum iterations is fixed to 20, and the algorithm stops if there is a variation of less than 10^{-2} in the cost function between two iterations. For all experiments, the parameters in the floodplains are bounded between a Strickler of 5 and $50 \text{ m}^{1/3} \cdot \text{s}^{-1}$. The background covariance matrix is fixed to $\mathbf{B} = 100\mathbf{I}_{N_{obs}}$ where $\mathbf{I}_{N_{obs}}$ is the identity matrix of size N_{obs} based on the

Strickler values found in the literature [Barnes, 1967; Chow, 1959]. For the observation covariance matrix, different values of $\mathbf{R}$ are tested in the twin experiment, and for the real experiment, the $\mathbf{R}$ value is fixed based on empirical estimates of the errors.

7.3.3.1 Twin experiment

First, the synthetic observations were assumed to be noise-free, setting the observation error covariance matrix to a low value, such as $\mathbf{R} = 10^{-10} \mathbf{I}_{N_{obs}}$. For the five calibrated subdomains, we first evaluated the convergence of the 3D-Var algorithm to the true values and tested the robustness of the method to the initial guess using 10 initial guesses. The algorithm starts from θ^b randomly drawn from a uniform distribution $\mathcal{U}(5, 50)$.

Figure 7.6 shows the evolution of the mean estimates of the Strickler values (for the 10 runs) in the five subdomains over iterations of the 3D-Var method.

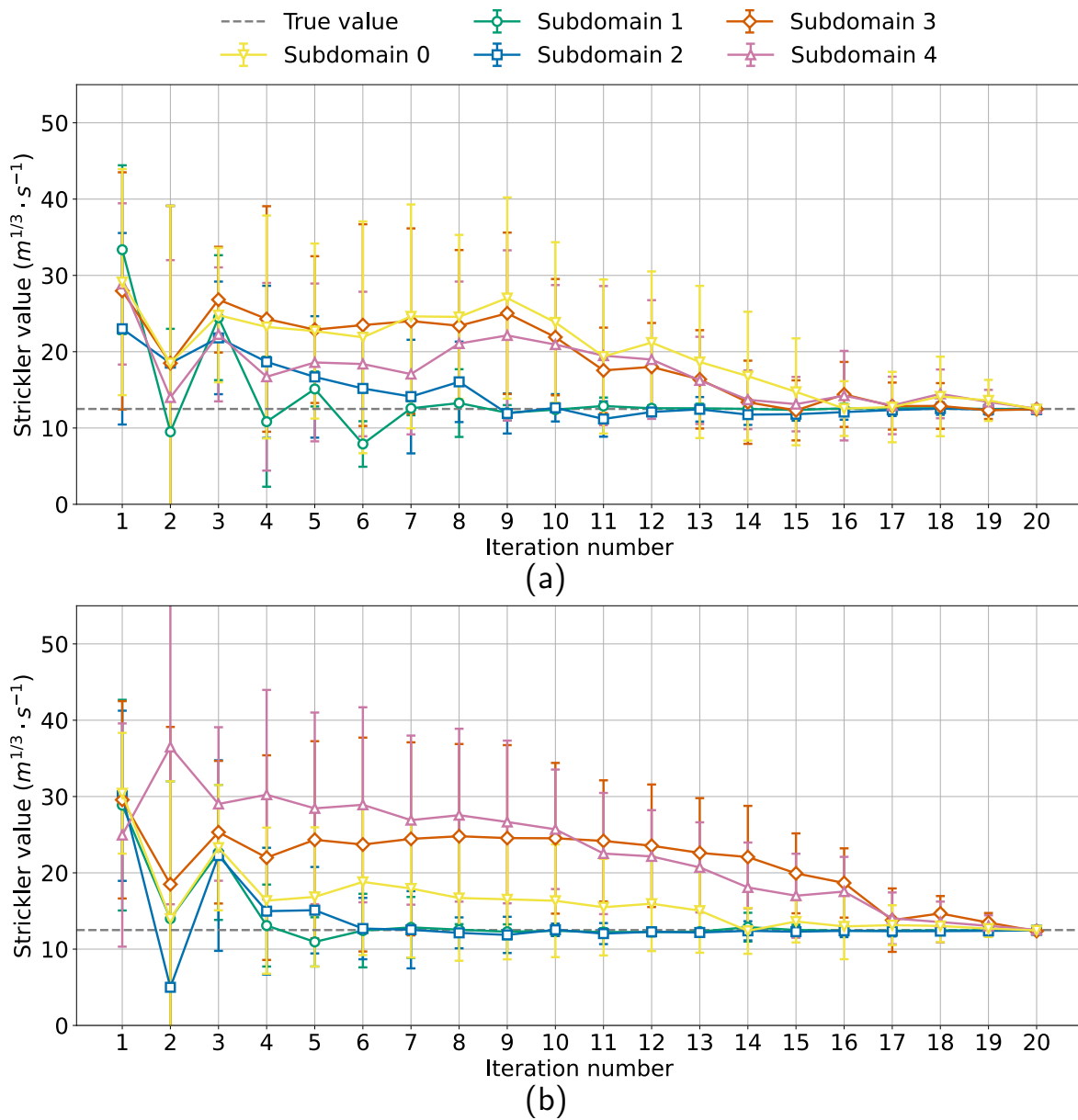


Figure 7.6: Mean estimates of the Strickler values over the 10 initializations in the five subdomains over iterations of the 3D-Var, for (a) configuration 1 with 100 observation locations and (b) configuration 2 with 20 observation locations. The error bars represent the standard deviation of the estimates on the 10 runs.

For both the configuration and the five subdomains, the algorithm converges to the true values with a varying number of iterations, depending on the subdomain. The subdomains 1 and 2 converged after a few iterations, while the other subdomains converged after more than 10 iterations to the true values. Subdomains 1 and 2 are the most influential on the water depth observations, explaining their faster convergence to the true values.

When the water depth observations are observed without noise in the floodplains, it is possible to calibrate the Strickler values in the larger zones (more influential zones on the simulated water depth field) of the study area. For configuration 1, the convergence is faster in subdomains 3 and 4 than in configuration 2, and inversely for subdomain 5 by a few iterations. These differences are mainly due to the distribution of the locations in the study area. However, overall, the second configuration is sufficient for calibrating the floodplain roughness with perfect observations.

7.3.3.2 Impact of observation errors

The DA framework is evaluated with observation errors using different levels of Gaussian noise on the synthetic water depth observations. The observation error covariance matrices writes as $\mathbf{R} = (\sigma^{obs})^2 \mathbf{I}_{N_{obs}}$, and noise levels varying between 10 cm and 1 m, such that $\sigma^{obs} \in \{0.1 \text{ m}; 0.2 \text{ m}; 0.5 \text{ m}; 1 \text{ m}\}$. These experiments aim to evaluate how close the parameters can be calibrated to the reference, given the observation errors, and determine the compromise between background knowledge and observation for varying noise levels.

Figures 7.7 and 7.8 present the evolution of the mean Strickler values over iterations for configuration 1 (100 observation locations) and configuration 2 (20 observation locations), respectively. For each noise level, 10 assimilation experiments are carried out with varying initializations θ^b . For all noise levels, the same 10 initial guesses are used.

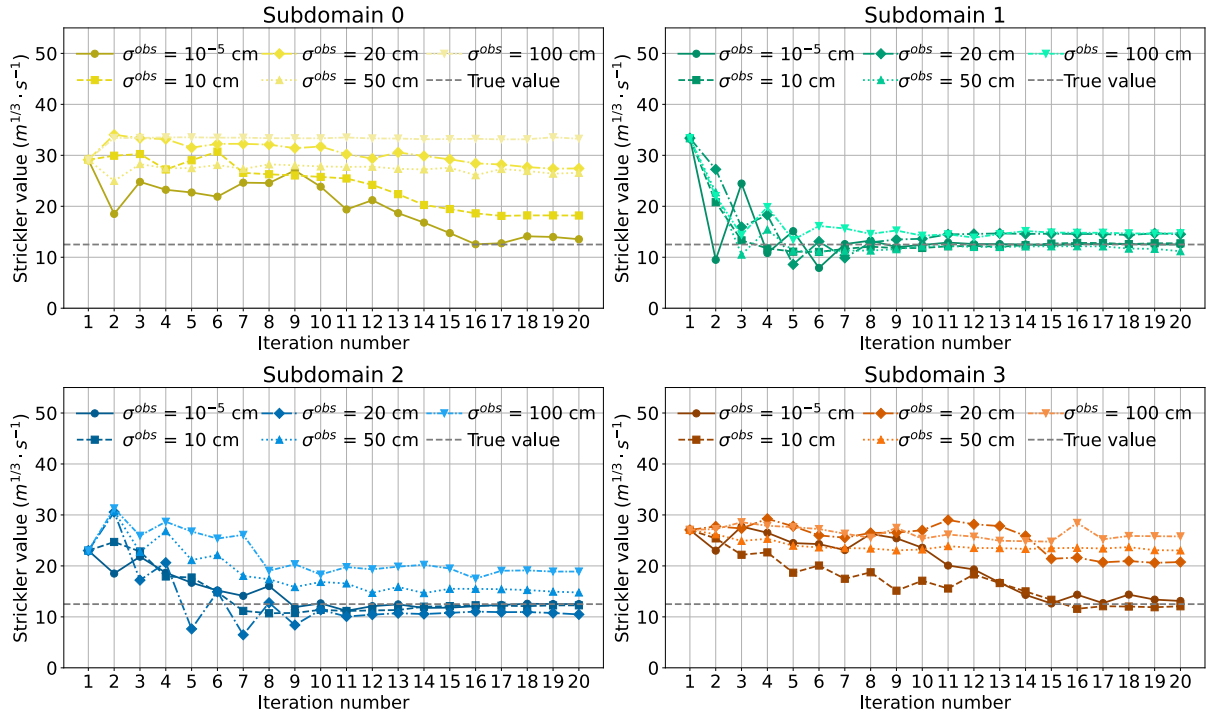


Figure 7.7: Evolution of the mean Strickler values (over the 10 runs) over 3D-Var iterations for configuration 1, using 100 observation locations. Results are shown for varying observation errors across subdomains 0 to 3.

For both spatial sampling configurations, the algorithm converges to the true value in each subdomain for perfect observations, as shown in Figure 7.6. Overall, for configu-

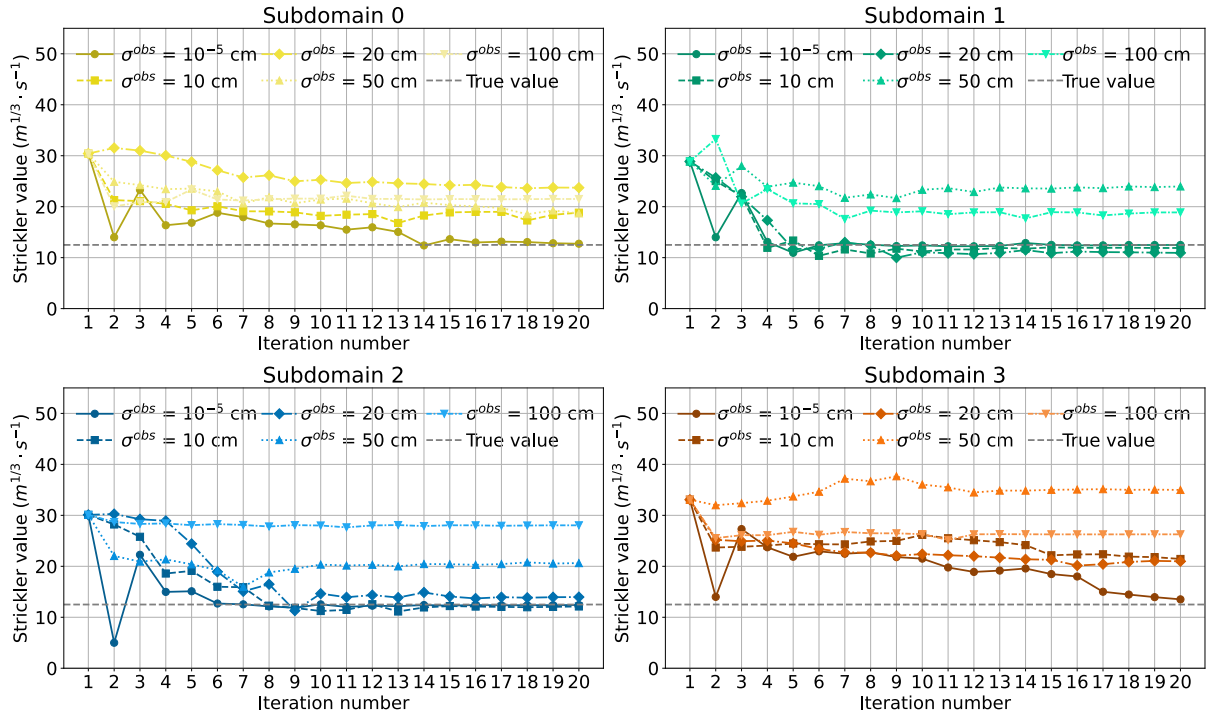


Figure 7.8: Evolution of the mean Strickler values (over the 10 runs) over 3D-Var iterations for configuration 2, using 20 observation locations. Results are shown for varying observation errors across subdomains 0 to 3.

ration 1 with more sampling locations, the estimates after 20 iterations are closer to the reference, even with higher levels of noise, particularly for subdomains 1 and 2. The better robustness to noise can be attributed to the larger number of observations, with more information. When $\mathbf{R}$ is large (high noise), the observation term in the cost function is weakened, but a large N_{obs} compensates by reinforcing the sensitivity of the cost function to the observation term. In subdomains 0 and 3, the estimates of the Strickler value are similar, meaning that the higher spatial density did not improve the characterization of the Strickler values.

Up to $\sigma^{obs} = 50$ cm, the Strickler estimates in subdomain 1 and 2 are near the true values for configuration 1 and up to $\sigma^{obs} = 20$ cm for configuration 2. For higher σ^{obs} , the estimates are further from the true values. For subdomains 0 and 3, the calibration of the Strickler values is more subject to the observation error, with estimated Strickler values higher than $12.5 m^{1/3} \cdot s^{-1}$ for all noise levels. However, it can be noted that for lower σ^{obs} , the estimates are generally closer to the reference than for higher σ^{obs} .

The results of the observation errors in the twin experiment suggest that both the spatial sampling and quality of the observations impact the performance of the DA framework for estimating the Strickler values in the sensitive subdomains of the study area. A high-density observation configuration (configuration 1) can partially compensate for larger observation errors, whereas a low-density configuration (configuration 2) requires higher accuracy to achieve reliable parameter estimates.

In the following, the assimilation of water depth observations for the two configurations is evaluated on the real experiment. The empirical estimates of observation errors are used. Therefore, for configuration 1, higher observation errors are present than in configuration 2, as explained in Section 7.3.1.

7.3.3.3 Real experiment

The initial values for the five subdomains were set to $\theta^b = [12.5; 12.5; 12.5; 12.5; 12.5]$, corresponding to the barycenter of the Strickler values used in operational EDF's flood studies for tall vegetation (see Table 2.4). In the real experiment, the observation errors should be specified in the 3D-Var cost function. The observation error covariance matrix is often evaluated empirically with in-situ data. The empirical estimates of the observation errors were quantified using available leash gauge data close to the observation locations (and reference simulations to avoid using outliers in the leash gauge data), although the leash gauges are also uncertain. For configuration 1, the estimated water depth observation errors ranged from 100 to 120 cm, while for configuration (with fewer observation points), the errors ranged from 50 to 70 cm. The difference in the two observation errors between the configurations is explained by the choice of the sampling strategy. In configuration 1, the points are chosen randomly within the study area, and for configuration 2 hydraulic expertise was used to select points in specific regions where the water depth estimates are more accurate for the FLEXTH method (regions of shallow slope and smooth flood maps).

Water depth observations derived from SAR imagery (see Section 7.3.1.2) were assimilated. Three noise levels were considered for each configuration:

- configuration 1: $\sigma^{obs} \in \{1.0 \text{ m}; 1.1 \text{ m}; 1.2 \text{ m}\}$
- configuration 2: $\sigma^{obs} \in \{0.5 \text{ m}; 0.6 \text{ m}; 0.7 \text{ m}\}$

Figure 7.9 reports the estimated Strickler over 3D-Var iterations in subdomains 0 to 3, for both configurations.

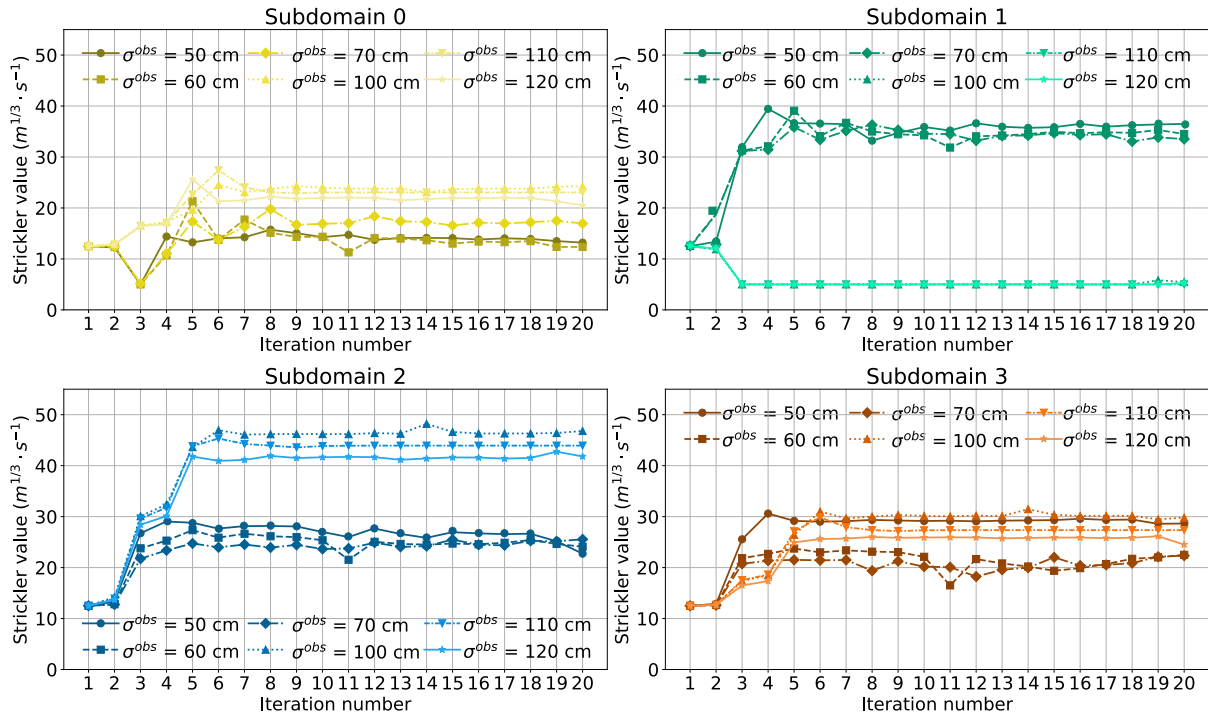


Figure 7.9: Estimated Strickler values over 3D-Var iterations in subdomains 0 to 3, using water depth observations from satellite imagery. Assimilation was performed with noise levels of $\sigma^{obs} \in \{1.0 \text{ m}; 1.1 \text{ m}; 1.2 \text{ m}\}$ for configuration 1 and $\sigma^{obs} \in \{0.5 \text{ m}; 0.6 \text{ m}; 0.7 \text{ m}\}$ for configuration 2.

In all subdomains, the assimilation converged quickly within a few iterations. The variation of the observation within each configuration did not significantly affect the

estimated Strickler values, except in subdomain 3, where the estimate with $\sigma^{obs} = 0.5$ m deviated from those at 0.6 and 0.7 m. Since the variation in observation errors had a limited impact on the results, we selected $\sigma^{obs} = 1.1$ m for configuration 1 and $\sigma^{obs} = 0.6$ m for configuration 2 for the rest of the analysis.

In subdomains 0 and 3, the same behaviors were observed for both configurations. However, in subdomains 1, configuration 1 converged to the lower bound of the parameter range, whereas configuration 2 converged to around $35 \text{ m}^{1/3} \cdot \text{s}^{-1}$. Subdomain 2 showed the opposite trend, with configuration 1 resulting in higher Strickler values than configuration 2. It highlights the impact of the observation location on the identification of Strickler values in 2D hydrodynamic models.

For subdomains 0, 2, and 3, the estimated Strickler values suggest that the parameter values should be higher compared to the initial guess. For the uncalibrated free-run experiment, the water depths are likely overestimated, and the water depth velocities are underestimated. For subdomain 1, the same behavior is observed for configuration 2, but for configuration 1, the estimated Strickler is lower than the background guess.

Figures 7.10 and 7.11 show the water depth differences between the assimilated model and the reference simulation (Strickler values corresponding to the barycenter of the ranges of Table 2.4) for both configurations, using observations from February 2 and 3, 2021.

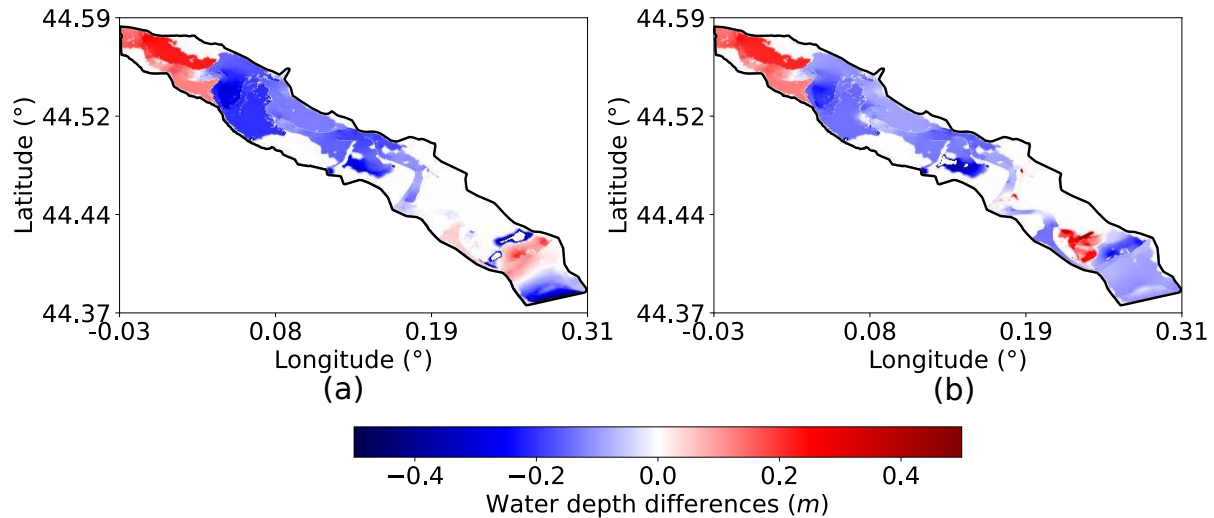


Figure 7.10: Water depth differences between the assimilated and reference simulations on February 2, 2021, at 17:56 UTC for (a) configuration 1 and (b) configuration 2.

For both acquisition times and configurations, the water depth of the reference simulation is mainly overestimated, especially in the central part of the study area, with water depth differences reaching up to 0.5 m. On February 2 (Figure 7.10), the assimilated water depths were similar for configurations 1 and 2, with slight discrepancies upstream. On February 3 (Figure 7.11), however, in the upstream zone, configuration 2 underestimated the water depth compared to the reference simulation. This can be explained by the low Strickler value estimated in subdomain 1, which reached the lower bound of the parameter range, indicating rougher terrain and leading to higher water depths in and around that subdomain (especially upstream because the flow is subcritical).

The estimated water depths and hence Strickler values represent a compromise between the background and the assimilated observations. To calibrate the model with greater accuracy, denser or more precise observations would be required. While a greater number of observations generally improves the results, selecting points with hydraulic expertise where the observation errors are constrained can lead to similar outputs in the assimilated water depths.

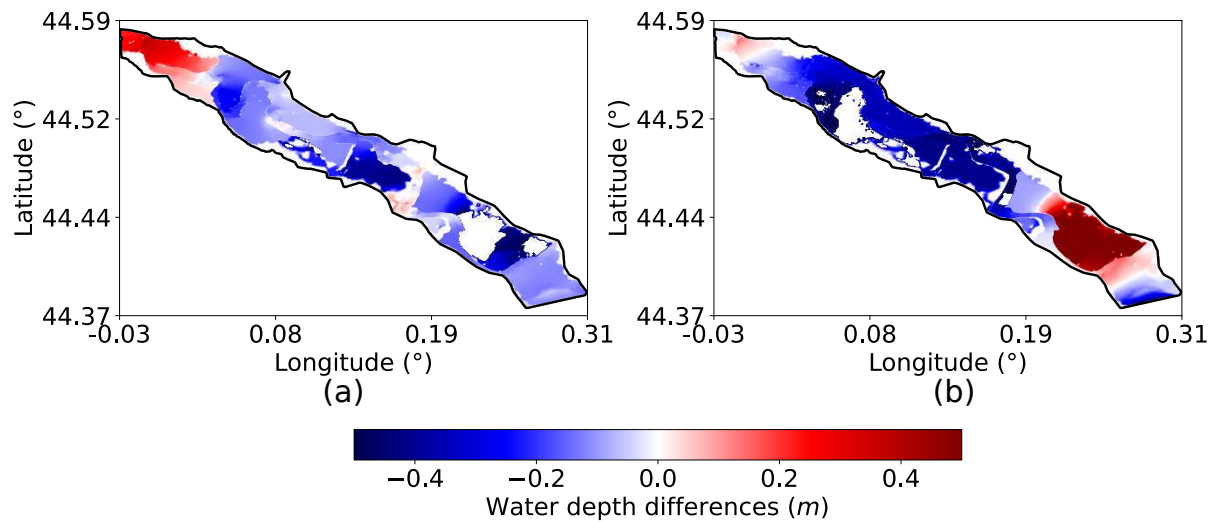


Figure 7.11: Water depth differences between the assimilated and reference simulations on February 3, 2021, at 17:48 UTC for (a) configuration 1 and (b) configuration 2.

7.4 Summary

Chapter 7 focused on calibrating floodplain roughness in 2D hydrodynamic models by assimilating water depth observations derived from satellite observations. First, the chapter presented the assimilation on a schematic test case using three DA techniques: the 3D-Var, the Ensemble Kalman Filter, and the Particle Filter. The schematic case study demonstrated that the 3D-Var algorithm was robust for parameter estimation. The other algorithms produced consistent results, but with higher computation needs. The schematic test case also highlighted the importance of observation locations and the impact of the observation errors on the accuracy of the estimated roughness coefficients.

Then, we tested the 3D-Var algorithm for assimilating water depth on a real case study on the Garonne River. The study evaluated the impact of the spatial distribution of water depth observations and observation errors on the calibration of the Strickler coefficients. The study was conducted both in a twin experiment setup and using water depth generated based on flood maps derived from satellite observations. First, we showed that the number of subdomains that can be calibrated can be reduced based on a sensitivity analysis. Then, we highlighted the impact of the observation errors and of the spatial observation density. In particular, we noticed a trade-off between the number of observation points and the accuracy. Selecting fewer but more accurate observation points based on hydraulic expertise in floodplains gave comparable results to configurations using a larger number of noisier observations.

This chapter demonstrated the possibility of calibrating spatially varying coefficients when high-density and high-precision water depth observations are available. However, the uncertainties in water depth observations extracted from satellite observations are still important and should be reduced to calibrate a model for future forecasting. In our case, the water depths observations are very uncertain, so DA only gives a compromise between the background knowledge and the observations. We showed that the compromise between the background knowledge and the observations led to a decrease in water depth in most of the study area. Compared to the assimilation of flood maps, observation errors are easier to understand and integrate in the DA framework.

Chapter 8

Conclusions and Perspectives

This thesis contributes to improving the quantification of roughness parameters in floodplains of 2D flood models using satellite observations. In this chapter, we highlight the main contributions of this work and provide new perspectives on flood model calibration using remote sensing data.

8.1 Conclusions

Research Question 1

What information can be derived from satellite observations to calibrate 2D flood models, and how accurately can it be quantified?

The present work began by identifying the potential of satellite observation to provide data for calibrating flood models, particularly in floodplains where in-situ data are scarce. The literature review conducted in Chapter 2 highlighted that SAR satellites were the most suitable for monitoring floods with missions such as Sentinel-1. Optical satellites were excluded due to frequent cloud cover during floods, and conventional nadir altimeters were dismissed as they provide only point-wise measurements. From SAR imagery, flood extent maps and water depth fields can be extracted.

SAR images consist of raw backscatter grids, which require specific processing steps for extracting usable hydraulic information. Chapter 2 reviewed the most widely used techniques for SAR image preprocessing, flood mapping, and water depth estimation. The review revealed a wide range of methodological options and tunable hyperparameters in the pipeline to extract flood maps or water depth fields. Therefore, in Chapter 4, we evaluated the influence of these processing strategies on the uncertainties of satellite-derived flood maps and water depth fields. This evaluation was conducted in two case studies: the Garonne River in France and the Shannon Basin in Ireland.

Flood maps are the most natural product derivable from SAR imagery. The study evaluated various flood mapping methods (thresholding, active contour, change detection, and supervised classification). While the general geometric structure of floods was reliably identified, significant errors remained around the flood boundaries and in vegetated and urban areas. It is caused by geometric distortions and measurement noise (as discussed in Chapter 2), which can lead to segmentation errors such as missed flooded zones, disconnected patches, or nonphysical holes in the flood extent. This study highlighted the sensitivity of flood map outputs to speckle filtering, resulting in variations of several square kilometers in the flooded area. Among the processing steps, the choice of flood mapping method had the most impact on the flood maps outputs. In the Garonne

case study, we demonstrated that the variability of the flooded area caused by processing can vary by up to 10-20 km² (preprocessing, flood mapping method, and hyperparameter tuning) exceeded the flooded area variations (e.g., 1 to 3 km²) caused by changes in floodplain roughness parameters within the 2D flood model (except when considering the channel roughness). This underlines that, given the uncertainty in flood map generation, only very sensitive models with broad parameter ranges can be meaningfully improved using such data for now.

By combining flood maps with Digital Elevation Models, water depth fields can be estimated. Using three different water depth estimation approaches, we found that the flood map inputs had a significant influence on the estimates. Then, errors in water depth estimation were largely driven by upstream uncertainties in flood map generation. Overall, water depth estimates from SAR imagery remain uncertain for operational uses with errors reaching up to 1 m. With this degree of uncertainty, we recommend incorporating the expertise of hydraulics experts, which is also pointed out in the methodology proposed by Hostache et al. [2009]. For instance, it can consist of selecting only water depth observations that are located in specific zones with higher confidence and reduced errors, characterized by low topography gradients and smooth delineation of the flood map.

This work underscores the importance of carefully selecting the processing workflow when using SAR images. Each stage, from preprocessing to method selection and hyperparameter tuning, can propagate errors that affect the final outputs. In many operational flood model calibration studies, flood maps are generated with a single fixed configuration. However, practitioners should account for uncertainty due to the processing to avoid calibrating models against poorly segmented flood maps or water depth estimates.

However, hydraulic modeling requires several steps, including spatial data preparation, model construction, parameter tuning, and results visualization. These tasks are time-consuming and sometimes not user-friendly, particularly when using command-line scripts or a sequence of independent tools. Thus, to support practitioners to account for the uncertainties in the generated outputs, a user-friendly graphic interface was developed. The Graphical User Interface enables the processing of SAR images, from downloading data to flood mapping and water depth estimation, with various configurations in a few clicks. This tool, alongside Jupyter notebooks and source code, facilitates the integration into hydraulic model calibration workflows.

In conclusion, these data can support the calibration and validation of hydrodynamic models in data-sparse regions or for very sensitive numerical models; however, they remain too uncertain for calibrating models with low sensitivity to parameter variations.

Research Question 2

How can the agreement between a satellite observation and a simulation be quantified?

The primary objective of this thesis is to use satellite observations to calibrate 2D flood models. In automated calibration strategies, the discrepancy between model outputs and observations is quantified using a mathematical tool known as *performance measure*. In Chapter 5, we evaluated how to compare flood models with satellite-derived flood maps, particularly those obtained from SAR imagery, which consist of dense binary images with distinct geometric textures.

Numerous performance measures have been proposed in the literature to quantify the mismatch between binary grids with spatial textures [Gilleland et al., 2009]. Therefore, we developed a step-by-step methodology to guide the selection of *performance measures* to compare observed and simulated flood maps, for purposes of model calibration or

validation.

The methodology begins with a correlation analysis to eliminate redundant performance measures. Then, a set of criteria is applied to determine whether a performance measure is sensitive to different types of errors, including position errors (translation and rotation errors), magnitude differences, and observation noise. The selected performance measures meeting the criteria are ranked based on their ability to identify consistent parameters under observation uncertainty.

The methodology was discussed numerically on a set of 24 performance measures on the 2021 flood event on the Garonne River in France. The flood model was parameterized with a spatially variable roughness field. This benchmark demonstrated the effectiveness of the methodology in identifying a reduced subset of performance measures. Thanks to the correlation analysis, we selected performance measures that satisfied the criteria and considered the best-ranked ones, identifying four reliable performance measures for comparing simulated and observed flood maps.

Using the selected performance measures, we presented a qualitative example illustrating how they can be used in operational calibration contexts. Using the simulated samples with spatially varying roughness, we identified for each of the performance measures the distribution of the roughness coefficients for the best-performing simulations. It highlights that the roughness coefficients were overestimated or underestimated in some subdomains, which impacts the simulated water depths and velocities. We also demonstrated that, depending on the selected performance measure, variations in flood front propagation could be observed, highlighting the importance of choosing an appropriate performance measure. Therefore, in operational flood studies, the choice of performance measures for comparing observed and simulated flood maps is crucial. This methodology can help practitioners calibrate model parameters using performance measures adapted to the properties of satellite flood maps. Although the methodology was numerically evaluated for the 2021 Garonne River flood event, it is designed to be generic and applicable to any flood model or flood event.

Research Question 3

Is it possible to calibrate floodplain roughness in 2D hydrodynamic models using uncertain satellite observations?

In response to Research Question 1, we showed that flood maps and water depth fields can be derived from SAR imagery, despite the uncertainties of these observations (Chapter 4). Chapter 5 addressed the challenge of selecting reliable *performance measures* for comparing observed and simulated flood maps. Building on this, all the necessary ingredients were available to infer the spatially varying floodplain roughness based on satellite observations, either using flood maps or water depth observations with DA. Chapters 6 and 7 explored whether satellite-derived observations can be used to calibrate floodplain roughness in 2D hydrodynamic models, through synthetic twin experiments and real case studies.

First, Chapter 6 focused on the assimilation of SAR-derived flood maps to estimate the spatially varying floodplain roughness field. The chapter proposed a two-phase strategy. First, a sensitivity analysis to reduce the number of parameters and limit the computational burden of the calibration phase in operational studies. Then, a variational algorithm (3D-Var) was applied as it was more adapted for using other performance measures than the weighted Euclidean distance and was shown to be the best strategy on our test case with limiting cost (in Chapter 7). The results of the twin experiments confirmed the feasibility of using flood maps for calibrating flood models when using ap-

propriate performance measures (identified in Chapter 5). However, applying the same strategy to real flood events is more challenging due to high uncertainties in flood map delineation. This highlighted the need for improved quantification of observation errors in flood maps. Currently, the uncertainties in satellite-derived flood maps are too large for reliable calibration of floodplain roughness across the parameter range typically used, even when extending the parameter range of the floodplain roughness commonly used in EDF flood studies. The flood maps can be used, but rather for calibrating more sensitive parameters, such as channel roughness or discharge in ungauged catchments. Indeed, when considering channel roughness as an additional uncertain parameter, it enhances the model's sensitivity, and the algorithm converges to more plausible values of roughness in the channel and floodplains. In conclusion, flood maps extracted from remote sensing data can help calibrate very sensitive models for now, and less sensitive models if we manage to improve flood map delineation and observation errors quantification. However, to support this conclusion, additional experiments should be carried out with noise on the synthetic observations, however, adding noise to flood maps and computing the associated observation errors for the performance measures is difficult.

While flood maps remain too uncertain for calibrating precisely the floodplains' roughness, Chapter 7 investigated the assimilation of another variable: water depth observations in the floodplains. In a schematic test case, we showed that the spatial location of observations is important and that selecting points in areas where water depths are more sensitive to roughness variations significantly improves calibration accuracy. We also highlighted that, depending on the number of observations, it was possible to calibrate the spatially variable roughness for variable noise levels on the water depth observations. Furthermore, by testing different DA frameworks, we demonstrated that 3D-Var was the most efficient parameter estimation algorithm for this test case with limited computational cost (despite not having an adjoint model). In the Garonne River case study, the same conclusions were underlined. With two spatial densities of observations, we demonstrated that additional points in the floodplains enhance the calibration of floodplain roughness, even for more uncertain data. However, selecting fewer points with hydraulic expertise reduces water depth errors and leads to comparable results during assimilation. When assimilating real observations, the compromise between the background and the observations led to an overestimation of water depths due to low Strickler values across most of the study area.

8.2 Perspectives

The contributions of this thesis highlight certain limitations, leading to the identification of future directions and unresolved questions, which are described below.

8.2.1 Reducing observation uncertainties and use of complementary data sources

The assimilation of satellite-derived flood maps and water depth fields from SAR imagery highlighted the difficulty in calibrating floodplain roughness due to the high uncertainties in the observations compared to the model's sensitivity to floodplain roughness. These uncertainties come from measurement errors, as well as the choice of processing method as shown in Chapter 4. Therefore, future work should prioritize reducing observation uncertainties and exploring complementary data sources to improve flood model calibration.

This thesis focused on using SAR imagery; however, other satellite sensors could also be used to extract hydraulic information. For instance, in the absence of cloud cover, optical imagery can be used, as it offers more interpretable data with more precise segmentation, thereby reducing observation uncertainties [Schumann et al., 2022]. Furthermore, the Sentinel-1 images used in the study have a spatial resolution of 10 m and a temporal resolution of 3 to 4 days. High spatial and temporal resolution sensors can be used when using optical data such as PlanetScope images, which have a spatial resolution of less than 3 m and a daily revisit frequency [Roy et al., 2021]. Furthermore, with the growing number of satellite monitoring floods, multi-sensor fusion could reduce uncertainty in flood map delineation [Zhao et al., 2024]. As for the water depth field, new satellites, such as the SWOT mission, could provide dense water height fields with observation errors below 10 cm when averaging over water areas above 1 km² [Biancamaria et al., 2016].

The thesis focused on using satellite observations, but other remote sensing technologies, such as drones, could be used, offering a closer look during flood events and providing near-real-time data. Nowadays, drones are often used in disaster response to monitor natural hazards, such as floods. In the future, during flood events, drones could be deployed within minutes above the study area to collect data for calibrating and validating hydrodynamic models or updating models in real-time [Karamuz et al., 2020]. Drones can generate high-resolution flood maps and carry LiDAR sensors to measure water surface elevation with centimetric precision. Additionally, leash gauges were not exploited in the thesis, but they can serve as complementary data to calibrate floodplains, especially with the development of crowd-sourced water levels [Dasgupta et al., 2022]. Based on the analysis of the importance of observation locations in floodplains for calibrating roughness, water depth gauges could be installed in floodplains at specific locations informed by the analysis of numerical models.

Flood maps and water depth fields were assimilated in this work, but future work should also explore the assimilation of velocity fields. In collaboration with the Laboratoire universitaire des sciences appliquées de Cherbourg (LUSAC), depth-averaged current speeds measured by ADCPs were assimilated in a tidal-stream resource assessment context. It highlighted the possibility of inferring the spatially variable Strickler values and parameters of the model to improve the resource assessment in the Alderney Race. The same methodology can be transferred to improve the quantification of model parameters in flood models. While ADCPs measurements are complicated during flood events, drone footage could be used to measure the surface velocity field using LSPIV. LSPIV measures surface water flow velocities by tracking the movement of texture in water flow images [Fujita et al., 1998]. With this technique, smartphone videos of flood events can also be

used to derive velocity fields as illustrated in Figure 8.1, with an application developed within EDF for computing velocities and discharge based on videos. Recently, velocity fields were also derived from satellite observations using LSPIV, with some satellite missions capturing short video clips around each image acquisition [Masafu and Williams, 2024].

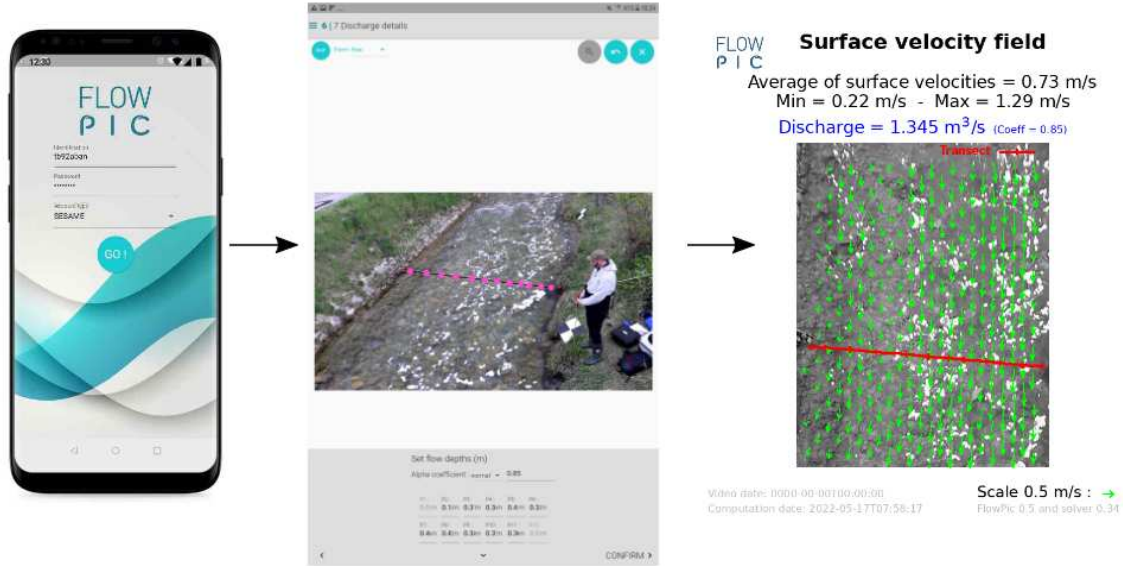


Figure 8.1: Example of a Large Scale Particle Image Velocimetry measure with FlowPic smartphone application (extracted from [Bodart, 2023]).

Finally, controlled experiments using physical models with floodplains could be conducted in laboratories to evaluate the calibration process with controlled measurements. In laboratory experiments, water depths can be measured at various locations in the floodplains, and drones can be used to derive the flood extent or velocity fields. In laboratory conditions, the quantification of observation errors, which was a key limitation identified in this thesis, is more controlled.

8.2.2 Data assimilation strategies and performance measure selection

In Chapter 5, we proposed a methodology for identifying global performance measures to compare modeled and observed flood maps using a single score. The methodology was applied to a benchmark case, but it is also transferable to any flood event or hydrodynamic model. Future works should apply this methodology on additional flood events, hydrodynamic models, satellite observations, and flood extraction methods to find general guidelines on which performance measures to use in which context (e.g., depending on catchment or flood characteristics). Furthermore, the methodology was developed for comparing observed flood maps to flood models, but in the future, the methodology could be adapted to other types of observations, such as water depth or velocity fields.

While the proposed methodology focused on comparing flood maps with global performance measures, other observations from satellite observations can be considered, and the observation operator in the 3D-Var cost function could be revised. For instance, Cooper et al. [2019] updated the water depth in a 2D numerical model using an observation operator that directly uses observed backscatter values of SAR imagery to avoid additional

steps such as flood mapping. Hostache et al. [2018] and Dasgupta et al. [2021b] assimilated probabilities of pixels of being flooded and Nguyen et al. [2022] assimilated the wet surface ratio in different subdomains. While flood maps and water depth fields were assimilated in this thesis, these observation operators are alternative options for calibrating 2D flood models.

However, a more promising direction is to consider that the information content between two similar flood maps is localized close to the edges of the flood extent in the flood front propagation zone. Other performance measures could be used to focus on these specific zones, thereby avoiding the dilution of informative data when using the entire flood map information. This approach was not initially considered due to the difficulty of precisely identifying flood edges, but it could be considered in the future with improved detection of flood maps or new sources of data, as specified in Section 8.2.1. In Rochoux et al. [2015] they assimilated with an EnKF the differences in coordinates between a simulated and observed fire front (with low measurement errors) as illustrated in Figure 8.2. These approaches can be used on flood fronts to match the simulated and the observed flood fronts by adjusting the roughness field.

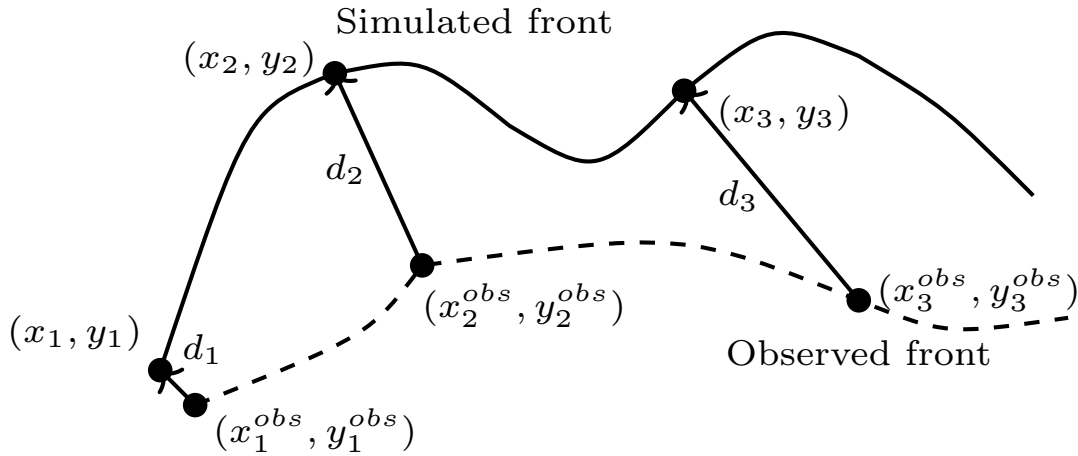


Figure 8.2: Differences between simulated and observed fire fronts (adapted from Rochoux et al. [2014]).

A promising research direction is to “move” along the intermediate states of the flood propagation front. Optimal transport provides an interesting framework to find the most efficient way to move mass (or probability) from a source distribution to a target distribution while minimizing some cost. For instance, when considering the Wasserstein distance (a key component of optimal transport), which measures the distance between two distributions instead of the Euclidean distance point to point, the analysis states between an observed and simulated water depth distribution on a cross-section is more physical as shown in Figure 8.3.

To reach the analysis state in Figure 8.3b), the data assimilation strategy should “move” between the probability distributions on the Wasserstein space. Most DA strategies were developed for Euclidean spaces. Now, to assimilate probability distributions (or water depth distributions along a cross-section) on a manifold equipped with Wasserstein distance, we should rather use algorithms adapted to manifolds, as the Manifold Extended Kalman Filter recently proposed in Le Ruz et al. [2025]. Ongoing work is being carried out to utilize data assimilation on the manifold of Wasserstein spaces to assimilate

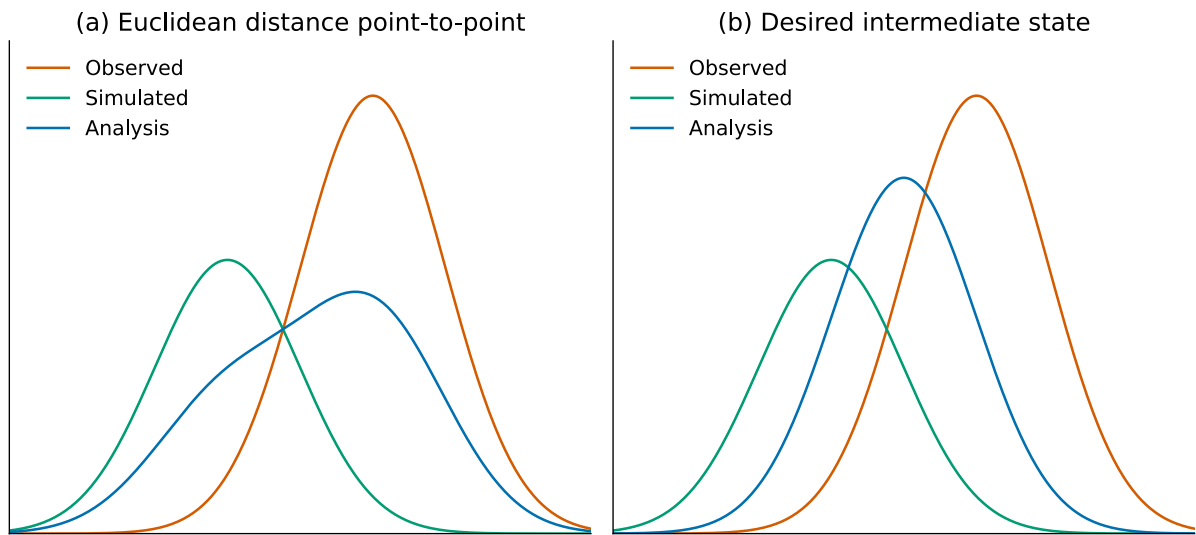


Figure 8.3: Example of intermediate states of water depth distribution on a cross-section with (a) Euclidean distance point to point, and (b) desired behavior.

water depth distributions. Then, the same work could be carried out using flood maps by comparing the flooded status of pixels along cross-sections. For interested readers, more details on this perspective, along with a description of the current limitations, are provided in Appendix A.2.

The Wasserstein space is based on Optimal Transport theory, but other possibilities are available in the literature for moving from a source distribution to a target distribution, such as flow matching approaches based on diffusion models [Lipman et al., 2022; Gat et al., 2024]. These methods aim to learn vector fields between the source and target distributions as visualized in Figure 8.4, often formulated as solutions to stochastic differential equations. New data assimilation frameworks based on diffusion processes have been developed recently [Rozet and Louppe, 2023a] and applied for estimating trajectories of dynamical systems using uncertain observations, for instance, on quasi geostrophic models [Rozet and Louppe, 2023b].

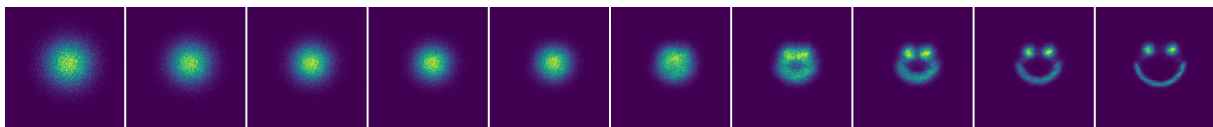


Figure 8.4: Example of trajectory between a source distribution (left) and target distribution (right) using flow matching.

List of Acronyms

- 1D** One-Dimensional. 30, 35
- 2D** Two-Dimensional. 17, 18, 23, 26, 27, 33, 35, 43, 45, 58–60, 75, 81, 89, 92, 133, 203, 204, 225–228
- 3D** Three-Dimensional. 23, 26, 30
- 3D-Var** Three-Dimensional Variational. 86, 87, 89, 92, 203, 204, 206, 220, 221, 224, 226, 229, 230, 232
- 4D-Var** Four-Dimensional Variational. 86–89
- ADCPs** Acoustic Doppler Current Profilers. 231
- ANOVA** ANalysis Of VARIances. 71, 72
- APIs** Application Programming Interfaces. 28, 130
- ASNR** Authority for Nuclear Safety and Radiation Protection. 17
- BFGS** Broyden–Fletcher–Goldfarb–Shanno. 91
- BLUE** Best Linear Unbiased Estimator. 62, 77–80, 87
- CLC** Corine Land Cover. 33, 34
- CNN** Convolutional Neural Networks. 54
- DA** Data Assimilation. 18, 60, 62, 64–66, 74–79, 83, 85, 86, 89, 92, 94, 133, 174, 175, 220, 222, 223, 226, 229, 230, 233
- DEM** Digital Elevation Model. 56–58, 218
- DoE** Design of Experiment. 67, 68, 72
- EDF** Électricité de France. 16, 17, 30, 33, 224
- EKF** Extended Kalman Filter. 79, 81, 88
- EMS** Emergency Mapping Service. 94, 126–128
- EnKF** Ensemble Kalman Filter. 79, 81–83, 85, 92, 204, 206, 233
- FwDET** Flood Water Depth Estimation Tool. 57, 217
- GSA** Global Sensitivity Analysis. 68–70, 73

- GUI** Graphical User Interface. 130, 131
- HPC** High Performance Computing. 92
- HSIC** Hilbert-Schmidt Independence Criterion. 73
- InSAR** SAR Interferometry. 45, 54–57
- KF** Kalman Filter. 62, 78–83, 88, 89, 92
- KNN** k-Nearest Neighbours. 54
- L-BFGS** Limited-memory Broyden–Fletcher–Goldfarb–Shanno. 91
- LARS** Least Angle Regression Stagewise. 72
- LHS** Latin Hypercube Sampling. 67, 68
- LiDAR** Light Detection And Ranging. 34, 37
- LNHE** National Laboratory for Hydraulics and Environment. 16, 27
- LOS** Line Of Sight. 43, 44, 57
- LSA** Local Sensitivity Analysis. 68, 69
- LSPIV** Large Scale Particle Image Velocimetry. 231, 232
- MAP** Maximum A Posteriori. 79, 86
- MCMC** Markov Chain Monte Carlo. 78
- MODIS** Moderate-Resolution Imaging Spectroradiometer. 37, 39
- MOISE** Modélisation des Inondations Externes. 17
- PCA** Principal Component Analysis. 72, 73
- PCE** Polynomial Chaos Expansion. 68, 72, 92, 219
- PF** Particle Filter. 62, 84, 85, 92, 204, 206
- PRCC** Partial Rank Correlation Coefficient. 71
- RCS** Radio Cross Section. 45
- RF** Random Forests. 54
- RMSE** Root Mean Square Error. 33, 57
- SA** Sensitivity Analysis. 36, 66, 68, 73, 92
- SAR** Synthetic Aperture Radar. 16, 21, 41, 42, 44–49, 52–58, 60, 93, 94, 126, 175, 203, 218, 224, 227–229, 231, 232
- SIR** Sequential Importance Resampling. 85

- SRRC** Standardized Rank Regression Coefficient. [71](#)
- SVM** Support Vector Machines. [54](#)
- SWE** Shallow Water Equations. [18](#), [21](#), [23](#), [26](#), [27](#), [29](#), [59](#)
- SWOT** Surface Water and Ocean Topography. [40](#), [231](#)
- T2D** TELEMAC-2D. [27](#), [29](#)
- TLM** Tangent Linear Model. [81](#), [91](#)
- UQ** Uncertainty Quantification. [18](#), [33](#), [59](#), [62](#), [64–66](#), [92](#)

Appendix A

Appendix

A.1 Complements to Chapter 6 on a schematic test case

Influence of 3D-Var Dissimilarity Measure on Flood Model Calibration with Satellite Data

Jean-Paul Travern^{1,2,*}, Cédric Goeury^{1,2}, Vito Bacchi¹, Fabrice Zaoui¹, Sébastien Boyaval^{2,3}

¹ National Laboratory for Hydraulics and Environment, EDF R&D, Chatou, France

² LHSV, Ecole des Ponts, EDF R&D, Chatou, France

³ MATHERIALS, Inria, Paris, France

* Corresponding author: jean-paul.travern@edf.fr

Abstract—Numerical models are essential in fluvial hydraulics for predicting flood dynamics and designing mitigation measures. These models must be calibrated based on observations to improve their forecast accuracy. Traditionally, this calibration relies on hydrometric station data, but it is now possible to use spatially distributed satellite observations. From a satellite observation (e.g., Synthetic Aperture Radar), it is possible to extract flood maps, corresponding to images of pixels classified as dry and flooded. This type of information is used in this study to calibrate a spatially distributed friction field of a TELEMAC-2D model. A twin experiment is conducted to generate synthetic satellite observations. The aim is to find the optimal friction field by comparing the synthetic data (flood maps) with the simulated states.

The 2D hydraulic model is automatically calibrated using a 3D-Variational data assimilation algorithm, which minimises a cost function by balancing the fit to observations with prior knowledge of the parameters (based on expert opinions, lab experiments, etc.). While conventional algorithms use L^2 distance to measure observation-model misfits, alternative dissimilarity measures offer potentially improved calibration. This study evaluates four dissimilarity measures — Euclidean distance, Measure of Fit, Procrustes Analysis, and Normalized Mutual Information — and assesses their impact on the calibration of the friction field.

This experiment has shown that some dissimilarity measures are better suited to calibrate the flood model, especially in floodplains.

Index Terms—Shallow Water Model, Parameter Estimation, Data Assimilation, Satellite Observation, openTELEMAC

I. INTRODUCTION

Numerical models, especially Two-Dimensional (2D) hydrodynamic models, play an essential role in simulating flood dynamics and informing decision-making processes in fluvial hydraulics. These models predict the spatial and temporal dynamics of a flood, which is crucial for designing effective flood mitigation measures.

To produce reliable forecasts, these models must be calibrated against real-world observations. Calibration involves adjusting the model parameters so that the model outputs align closely with the observed data. In flood modeling, the main parameter to calibrate is often the uncertain friction field [1], which affects flow resistance and consequently modeled velocities and water depths. In most engineering studies, the friction field is often calibrated using hydrometric stations along the channel. However, these data points are often sparse and may not provide comprehensive coverage, especially in the floodplains. The

lack of detailed and spatially distributed data has historically posed a significant challenge to the calibration of hydraulic models, especially in floodplains. However, in recent years, there has been an increasing availability of spatially distributed observations thanks to satellites. For instance, Synthetic Aperture Radar (SAR) satellites can detect flood events regardless of weather conditions and provide high spatial resolution images (e.g., 10 meters for Sentinel-1). This capability allows for the generation of detailed flood maps consisting of images, i.e., a grid of pixels (also called rasters) classified as flooded or dry, as illustrated in Figure 1.

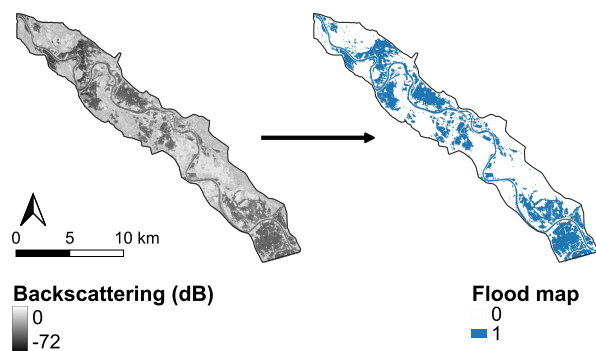


Fig. 1: Example of Sentinel-1 acquisition on February 2, at 5:55 p.m. on the Garonne River in France, and extracted flood map.

In this context, this work discusses the impact on the calibration of the choice of mathematical tool to compare observed flood maps and numerical simulations of a flood event.

Automatic calibration typically involves an algorithm that minimizes the discrepancy between model outputs and observations. Most optimization problems use the weighted L^2 distance to compute this discrepancy by comparing data points to points. However, in the case of flood images, this approach does not take into account the structure of the image and does not handle position errors [2]. Then, alternative dissimilarity measures for comparing spatially distributed fields can provide different insights and potentially improve the calibration of the model.

In this work, the friction field is automatically calibrated with a 3D-Variational (3D-VAR) data assimilation algorithm. This method aims to minimize a cost function that balances prior knowledge of parameter values (from lab experiments, expert opinions, etc.) with the misfit between model outputs and observations. The latter, referred to as dissimilarity measures, can be chosen from various possibilities to compare images [2, 3]. In this work, a few dissimilarity measures from the literature are selected, comparing flood maps on a point-to-point basis or based on their “geometrical” properties.

This paper aims to quantify the influence of the chosen dissimilarity measure on the calibration results. A purely numerical experiment is conducted by using generated observations from a twin experiment. The twin experiment consists of a schematic test case for one flood event to estimate an uncertain spatially distributed friction field based on one observed flood map.

Section II sets the parameter estimation problem, the data assimilation framework, and the choice of dissimilarity measure. Section III presents the test case for a schematic flood event in a twin experiment and the software tools used for the implementation. Then, Section IV describes and discusses the numerical results of the comparison of the use of the dissimilarity measures for the test case. Finally, Section V presents the conclusions and perspectives of this work.

II. PROBLEM SETTING

Parameter estimation consists of evaluating the input data of a problem from its “solution”. The aim of the paper is to estimate the spatially distributed friction field of a flood numerical model through a comparison between observed (generated with a twin experiment here) and simulated flood maps. In this section, the hydrodynamic model, the automatic calibration algorithm, and the dissimilarity measures are introduced.

A. Hydrodynamic model

Two-dimensional simulations of flood events are often computed as a solution of the 2D Shallow-Water-Equations (SWEs). The SWEs write in a Cartesian coordinate system where gravity acts uniform as $-g\vec{e}_z$ as:

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{u}) = 0 \quad (1)$$

$$\frac{\partial (h\mathbf{u})}{\partial t} + \nabla \cdot (h\mathbf{u} \otimes \mathbf{u}) = -g\nabla\eta + \nabla \cdot (h\nu_e \nabla \mathbf{u}) - \frac{\boldsymbol{\tau}_b}{\rho}, \quad (2)$$

where the water depth $h \geq 0$ and $\mathbf{u} = u\vec{e}_x + v\vec{e}_y$ are unknown variables function of x, y and time $t \in [0, T]$. We denote $\eta = h + z_b$ the free surface elevation where $z_b(x, y)$ is the bottom elevation, $\nu_e > 0$ a constant viscosity and $\boldsymbol{\tau}_b = \tau_{bx}\vec{e}_x + \tau_{by}\vec{e}_y$ the bed shear stress.

In this work, the two components of the bed shear stress are computed with the Strickler formula [4] as:

$$\begin{cases} \tau_{bx} = \frac{\rho \cdot g \cdot u}{h^{1/3} \cdot K_s^2} \sqrt{u^2 + v^2} \\ \tau_{by} = \frac{\rho \cdot g \cdot v}{h^{1/3} \cdot K_s^2} \sqrt{u^2 + v^2} \end{cases}, \quad (3)$$

where the spatially distributed Strickler coefficient $K_s(x, y)$ needs calibrating.

Equations (1-2) are solved with TELEMAC-2D (T2D), a module of the open-source hydro-informatics system openTELEMAC (www.opentelemac.org) [5]. The modeling is based on a finite element approach on an unstructured triangular mesh. The N edge-by-edge and the Positive Streamwise Invariant distributed schemes compute the advection of velocity and water depths, respectively. As for the wetting and drying of the grid elements (essential for floodplains floodings), it is considered through a correction of the free surface gradient [5].

B. Automatic calibration algorithm

In this paper, the calibration is performed using Data Assimilation (DA), which is a sub-class of inverse problems. DA aims at estimating uncertain states or parameters of a physical system, balancing the discrepancy between observations and model outputs, and a prior knowledge of the unknowns. In this work, we specifically use a variational approach called 3D-VAR, often used for parametric estimation problems [6, 7]. The goal of the algorithm is to evaluate the unknowns by minimising a cost function quantifying the misfit between model predictions and measurements and the prior knowledge of the parameters. The traditional variational data assimilation cost function writes as follows:

$$\begin{cases} J(\boldsymbol{\theta}) = J_b(\boldsymbol{\theta}) + J_o(\boldsymbol{\theta}) \\ J_b(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^b)^T \mathbf{B}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^b) \\ J_o(\boldsymbol{\theta}) = \frac{1}{2}(\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\boldsymbol{\theta})))^T \mathbf{R}^{-1}(\mathbf{y}^{obs} - \mathcal{H}(\mathcal{M}(\boldsymbol{\theta}))) \end{cases}, \quad (4)$$

where $\mathbf{y}^{obs}$ is the observation, $\mathcal{M}$ the numerical model, $\boldsymbol{\theta}$ the unknown parameters to calibrate, $\boldsymbol{\theta}^b$ a background knowledge of the parameters, and $\mathcal{H}$ is a *smooth* observation operator from the simulation to the observation space (preserving the smoothness of $\mathcal{M}$ as a function of $\boldsymbol{\theta}$). Background and observation errors are considered with error covariance matrices denoted $\mathbf{B}$ and $\mathbf{R}$ often considered as diagonal matrices. In this work, the observation $\mathbf{y}^{obs}$ shall represent a flood map, and $\mathcal{H}$ will transform a simulated water depth field into a flood map within a smooth manifold. This transformation will be described in Section III-A4.

For the traditional cost function in (4), Euclidean distances weighted by matrices $\mathbf{B}$ and $\mathbf{R}$ are used. In fact, the compromise between J_b and J_o can be written more generally as:

$$J(\boldsymbol{\theta}) = \omega_b d_b(\boldsymbol{\theta}, \boldsymbol{\theta}^b)^2 + d_o(\mathbf{y}^{obs}, \mathcal{H}(\mathcal{M}(\boldsymbol{\theta})))^2, \quad (5)$$

where d_b and d_o are distances to specify, and ω_b is a weight balancing the background knowledge and the observation. In this work, d_b is fixed as a usual L^2 distance, and the aim is to study the influence on the calibration of the choice of the dissimilarity measure d_o , supposedly piecewise differentiable on a flood-map manifold.

The optimal θ is found by minimizing J . This minimization is generally carried out using a gradient descent method such as L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno Bound-Constrained) [8]. The L-BFGS-B algorithm is based on the Newton method, using finite differences to compute an approximate gradient of $J(\theta)$. The procedure is iterative, and the algorithm explores all the directions of the parameters to determine how to update the parameter for the next iteration. The background knowledge θ^b is the starting point of the minimisation of the cost function.

C. Dissimilarity measures

This section aims to describe briefly the different dissimilarity measures d_o used in this work in (5). The usual Euclidean distance is tested, as well as three other dissimilarity measures, namely the measure of fit, Procrustes analysis, and Normalized Mutual Information. They are chosen from the literature to have measures based on different principles (pixel-to-pixel, shape alignment, information theory). Methodologies exist in the literature to choose robust dissimilarity measures between simulated and observed flood maps [9, 10]. Euclidean distance is usually used to correct errors in amplitude, while Procrustes analysis or Normalized Mutual Information may also be able to correct errors in the position of the flood.

Flood maps are images considered as a grid of pixels of size $N_x \times N_y$. In the following, the simulated and observed flood maps are denoted as $O, S \in \{0, 1\}^{N_x \times N_y}$ (i.e, 0 for a dry pixel, and 1 for a flooded pixel). Then, in the following $y^{obs} = O$ and $\mathcal{H}(\mathcal{M}(\theta)) = S(\theta)$.

1) *Euclidean distance*: In the most common case, the observed and simulated values are compared with an Euclidean distance two-by-two, so that:

$$d_e(O, S) = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (O_{i,j} - S_{i,j})^2}. \quad (6)$$

2) *Measure of fit*: Another commonly used measure to calibrate flood models is the measure of fit [3]. This measure is based on comparing the number of pixels that match or do not match between O and S . The four possible outcomes when comparing pixel values are listed in Table I.

	OBSERVED FLOODED	OBSERVED DRY
SIMULATED FLOODED	TP	FP
SIMULATED DRY	FN	TN

TABLE I: Classification of pixels between O and S .

For instance, TP represents the number of pixels classified as flooded in both O and S , so that $TP = \sum_{i=1}^N \sum_{j=1}^N \mathbb{1}(O_{i,j} = 1 \wedge S_{i,j} = 1)$. The measure of fit is computed as follows:

$$F^4(O, S) = \frac{TP - FP}{TP + FP + FN} \quad (7)$$

3) *Procrustes analysis*: Procrustes analysis [11] optimally transforms two point sets (translations, rotations, scaling), to minimize their differences. O and S are standardised into two matrices A_1 and A_2 (centered around their origins and $Tr(A_i \cdot (A_i)^T) = 1$ with $i \in \{1; 2\}$). Then, the Singular Value Decomposition is used to find an optimal transformation minimising (8):

$$P^2 = d_e(A_1 - cA_2Q), \quad (8)$$

where c is a scaling factor and Q an orthogonal matrix.

4) *Normalized Mutual Information (NMI)*: The NMI [12] quantifies the amount of information shared between two set of points. It is defined as:

$$NMI(O, S) = \frac{H(O) + H(S)}{H(O, S)} \quad (9)$$

where $H(O) = -\sum_{i=1}^{nc} p_i \log(p_i)$ is the Shannon entropy of matrix O , with nc the number of classes (2 here with flooded and dry pixels) and p_i the probability of class i . The joint entropy is defined as $H(O, S) = -\sum_{i=1}^{nc} \sum_{j=1}^{nc} p_{i,j} \log(p_{i,j})$ with $p_{i,j}$ the joint probability between classes i in O and j in S .

F^4 , and NMI vary between 0 and 1, and equal 1 when O and S are identical. However, the cost function J , should be minimized. Then, for these dissimilarity measures, the values are translated so that they equal 0 when O and S are identical. Then for the measure of fit, $d_o(O, S) = 1 - F^4(O, S)$, and similarly for NMI .

III. TEST CASE AND SOFTWARE TOOLS

The experiments are purely numerical, with the model $\mathcal{M}$ operating as a twin experiment. First, the test case consisting of a schematic river and flood event is described. Then, more details are given on the vector θ to calibrate, as well as describing the process to generate synthetic data and how $\mathcal{H}$ transforms simulated water depths into flood maps. Following this, an overview of the software tools is provided.

A. Definition of the test case

1) *Geometry*: A rectangular geometry is considered, with a 100 m wide meandering channel in the x direction and floodplains on both sides (see Figure 2). The topography is initialized such that the channel begins to overflow at the center of the domain, located at $x = 5,000$ m. This setup is designed to minimize sensitivity to boundary conditions. The channel depth decreases linearly from 2 m at $x = 0$ m to 1 m at $x = 5,000$ m, and then returns to 2 m at $x = 10,000$ m. On the floodplain the slope is fixed to

0.002 m/m. The geometry was already used in a previous study (see [13] for more details), and we built on these efforts as regards the construction of the model.

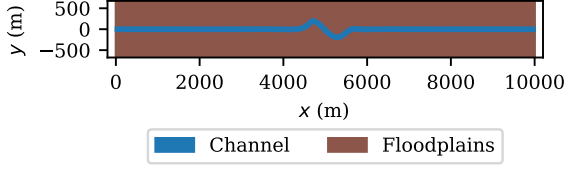


Fig. 2: Schematic meandering river with floodplains.

The geometry is similar to that in [13], but the mesh in the present study is coarser to reduce the computational cost. The T2D model uses a simplicial mesh composed of $\mathcal{N} = 7979$ nodes, with edge sizes varying from 20 m in the channel to 50 m in the floodplains.

2) *Initial conditions, and boundary conditions:* The model is initialized at $t = 0$ with a water depth of 1 m within the channel. A schematic time-varying discharge is considered upstream at $x = 0$, with a flow rate varying linearly from 100 to $450 \text{ m}^3 \cdot \text{s}^{-1}$ as shown in Figure 3.

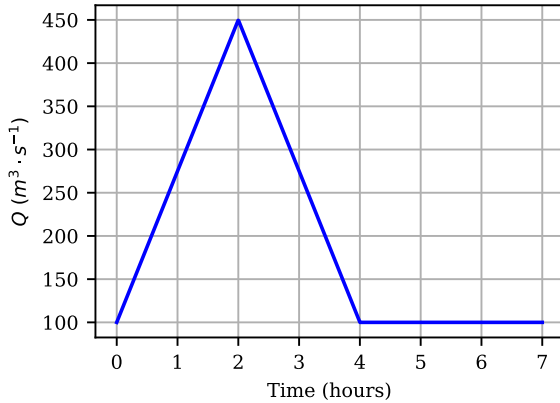


Fig. 3: Upstream flood hydrograph during the simulation.

A rating curve linking water depths to discharge is imposed downstream with water depths between 1 m and 2.6 m as a function of Q varying between $100 \text{ m}^3 \cdot \text{s}^{-1}$ and $450 \text{ m}^3 \cdot \text{s}^{-1}$ respectively. The total simulation time is set to 7 hours.

3) *Friction zoning:* The friction parameter to calibrate is spatially distributed. It is considered to be piecewise constant in the simulation domain, with a uniform value over subdomains. In this test case, three subdomains are considered, i.e. the channel, the left floodplain, and the right floodplain as illustrated in Figure 4.

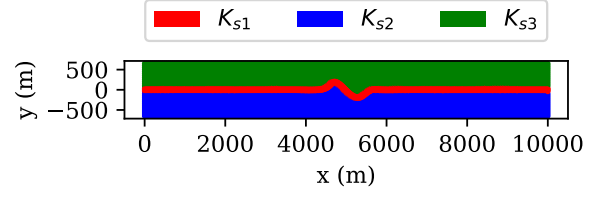


Fig. 4: Roughness subdomains for the experiment.

Then, the control vector of the parameter is $\theta = [K_{s1}, K_{s2}, K_{s3}]$.

4) *Observation operator and generated synthetic data:* The output of the simulation is a water depth field. However, the goal is to compare flood maps, then the water depth field should be transformed into a flood map. We denote $H_{i,j} \in \mathbb{R}^{N_x \times N_y}$ a matrix where $H_{i,j}$ equals the value of the finite element water depth at the center of the pixel (i, j) . Then, the observation operator transforms a simulated water depth field into a flood map as:

$$[\mathcal{H}(\mathcal{M}(\theta))]_{i,j} = S(\theta)_{i,j} = \begin{cases} 1 & \text{if } H_{i,j} > 0.01 \text{ m} \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

for the (i, j) -th component of the observation operator.

The experiment is purely numerical, so synthetic data must be generated. Then, one observation $O = \mathcal{H}(\mathcal{M}(\theta^t))$, with $\theta^t = [30, 20, 10]$ representing the true parameter values is generated at $t = 3 \text{ hours}$ of simulation. Figure 5 visualises the functioning of the observation operator to generate the synthetic data with θ^t .

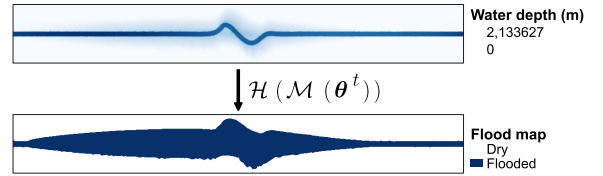


Fig. 5: Example of application of the observation operator with the parameter truth values at $t = 3 \text{ hours}$.

B. Software tools

In this work, the TELEMAC-2D solver (version V8P5) is coupled with TelApy, a module of the openTELEMAC system, and ADAO, a Python Data Assimilation library (<https://pypi.org/project/adao/>). These modules are presented briefly.

1) *TelApy:* TelApy is a module distributed with the openTELEMAC system. It is a practical module that wraps in Python the Fortran APIs of openTELEMAC [14]. It provides the practitioner the possibility to control the simulations while they are running. Then, it is particularly useful in the case of parameter optimization where the parameters are changed iteratively. It also enables the user

to retrieve the simulation states during the simulation, so they can be compared with the observations. The TelApy component allows the creation of optimization chains by coupling it with optimization algorithms.

2) *ADAO*: The automatic calibration is carried out thanks to a Data Assimilation library called ADAO [15]. ADAO provides various optimization and data assimilation algorithms, including the 3D-Var algorithm used in this work. In this work, the ADAO module is coupled to TelApy to retrieve the numerical model states and test new parameters.

IV. RESULTS AND DISCUSSION

After presenting the test case and the software tools, the different dissimilarity measures are tested to calibrate the model. The influence of the prior guess on the parameter on the calibration is also evaluated.

In this work, we consider a calibration in which the observation is considered almost perfect and the prior guess on the parameter is very uncertain. The weight ω_b is set arbitrarily to 10^{-12} , making the role of d_o in (4) significantly more prominent than that of d_b .

The prior guess for the parameter should also be specified. Initially, the friction parameter is set to $\theta^b = [35; 12; 18]$, but the impact of the initial state on the calibration results with different d_o is investigated in Section IV-B. In the L-BFGS-B algorithm, the values of $\theta = [K_{s1}, K_{s2}, K_{s3}]$ are constrained, such as $20 \leq K_{s1} \leq 40$, and $5 \leq K_{s2}, K_{s3} \leq 50$.

For computing the gradient $J(\theta)$ in the 3D-Var algorithm, the differential increment is set to 10^{-3} . The algorithm stops if the number of iterations exceeds 35 or when the cost function decreases less than 10^{-4} .

A. Influence of d_o on parameter estimation

The four dissimilarity measures are tested with the previous numerical considerations. In Figure 6, the values of the three parameters of θ are plotted at each iteration. An iteration corresponds to the exploration of the parameter space in every direction (here three directions), then θ is updated. In the case of three parameters, it corresponds to four TELEMAC-2D simulation runs. On an Intel Core i7-11850H @ 2.50GHz processor, each simulation run takes 36 seconds.

For the four dissimilarity measures, the friction parameter for the channel is well calibrated after a few iterations (5 to 6 iterations). For all dissimilarity measures, the channel Strickler value converges quickly, since it is the most influent parameter on the state of the pixels in the flood maps. For the different dissimilarity measures, the calibration of the floodplains show different behaviors. The dissimilarity measures are less sensitive to changes in floodplain friction values. Indeed, only the calibration with *NMI* reaches the true values of the generated observations ($K_{s2} = 20$ and $K_{s1} = 10$) at the end of the 35 iterations.

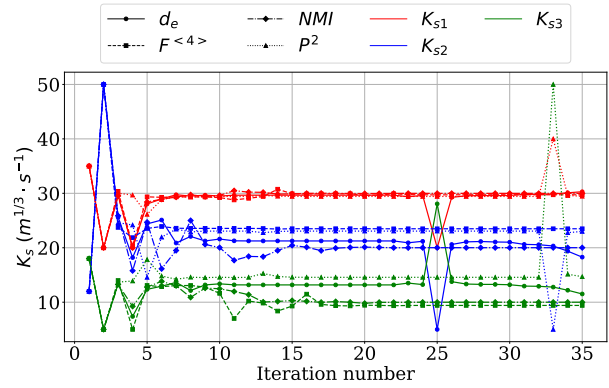


Fig. 6: Automatic calibration of θ for the four dissimilarity measures (circles for d_e (Euclidean distance), diamonds for *NMI* (Normalised Mutual Information), squares for $F^{<4>}$ (Measure of fit), and triangles for P^2 (Procrustes analysis)) for $\theta^b = [35; 12; 18]$.

For the three other dissimilarity measures, the automatically calibrated parameters on the floodplains are higher or lower than the expected values, whether for K_{s2} or K_{s3} . Thus, in this configuration, using *NMI* as the dissimilarity measure seems to give the closest calibration to the reference. It seems that the calibration with the Euclidean distance is not yet fully converged, and it may converge to the true values of the parameters for more iterations. However, one may prefer using dissimilarity measures that converge quickly. At iteration 32 with Procrustes analysis, the estimated values of the parameters seem to be outliers that may be explained by the variational approach testing a distant value to check if the algorithm is not stuck in a local minimum.

B. Influence of the initial state θ^b

In the previous Section IV-A, only one prior guess on θ^b serving as the initial parameter of the descent of 3D-Var is considered. Thus, other initial prior guesses on θ^b should be tested to confirm that using *NMI* yields better results to calibrate this model when comparing flood maps.

In Figures 7 and 8, two other initial guesses for the parameters are considered: $\theta^b = [33, 22, 8]$ and $\theta^b = [27, 25, 13]$. These initial guesses are visualized as they show the different behaviors of the calibration process. However, more exhaustive tests were conducted by sampling θ^b with the same bounds as for the L-BFGS-B algorithm ($\theta^b \in [\mathcal{U}(20; 40), \mathcal{U}(5; 50), \mathcal{U}(5; 50)]$). From a physical point of view, more restrictive constraints could be included by considering that the Strickler value in the floodplain should be smaller than for the main channel, but it was not considered here.

In Figure 7, for an initial guess $\theta^b = [33, 22, 8]$, the Euclidean distance did not reach the true values even for the channel (it stops after 23 iterations, because the

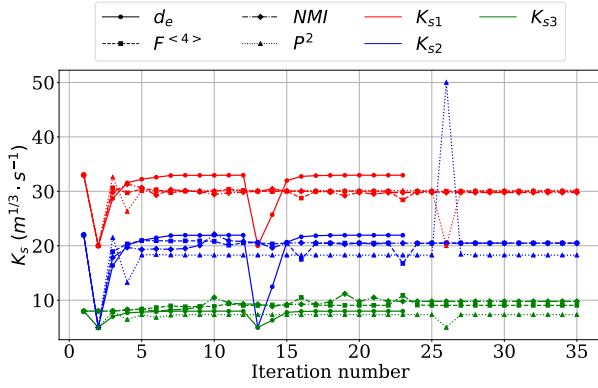


Fig. 7: Automatic calibration of θ for the four dissimilarity measures (circles for d_e (Euclidean distance), diamonds for NMI (Normalised Mutual Information), squares for $F^{<4>}$ (Measure of fit), and triangles for P^2 (Procrustes analysis)) for $\theta^b = [33, 22, 8]$.

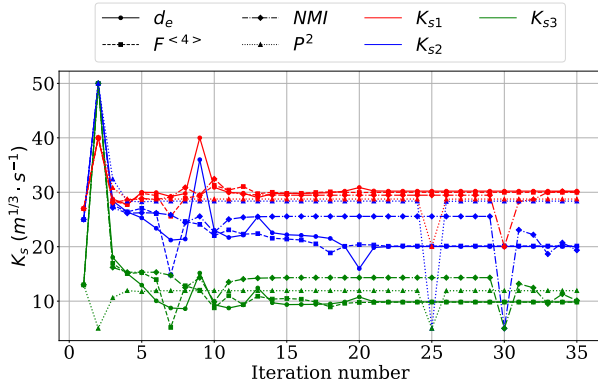


Fig. 8: Automatic calibration of θ for the four dissimilarity measures (circles for d_e (Euclidean distance), diamonds for NMI (Normalised Mutual Information), squares for $F^{<4>}$ (Measure of fit), and triangles for P^2 (Procrustes analysis)) for $\theta^b = [27, 25, 13]$.

stopping criterion is reached), while all the other measures are well calibrated for the channel. For the floodplains, again P^2 fails to reach the expected values. However, the behavior of the calibration with F^4 changes with that new initial guess. Indeed, while F^4 did not converge towards the true values for $\theta^b = [35; 12; 18]$, it did for $\theta^b = [33, 22, 8]$. Depending on the initialization, the cost function may get stuck in a local minimum.

Again, for the initial guess $\theta^b = [27, 25, 13]$, P^2 does not calibrate the model well. For the three other dissimilarity measures, they have reached the true values. However, for that initialisation, using NMI required longer to converge to the true values than for the previous initialisations, and it seems that 35 iterations is a little too small for a full convergence.

While the choice of dissimilarity measure is important for calibrating the model, the initial guess of the parameters also plays a role. From these three Figures, it seems that some dissimilarity measures are better suited for calibrating the model, regardless of the initial guess for the parameter. Other initial states of the friction values were tested. Although the exhaustive results are not visualized in the paper, using NMI seemed more robust for calibrating this model.

To understand the observed behavior, the variation of the dissimilarity measure is evaluated as a function of the variation of the three components of θ . Coupled variations of the parameters (two at a time) are tested by fixing one of the three parameters of θ . When two parameters are modified within their range of values (bounds defined previously), the third parameter is fixed to its true value ($K_{s1} = 30$, $K_{s2} = 20$, and $K_{s3} = 10$).

In Figures 9 and 10, the coupled variations of the channel and the floodplains friction values show that the four dissimilarity measures have a global minimum. However, it seems more difficult to find the global minimum for P^2 when exploring the parameter space, due to smaller gradients (especially for the coupled variations of K_{s1} and K_{s3}). Because of this smaller gradient, it may be more difficult to converge to the expected friction values on the floodplains. For the Euclidean distance and for $1 - NMI$ the global minimum is clearly identified and corresponds to the true values. Then, this analysis does not explain the discrepancy between the automatically calibrated values and the true values for the Euclidean distance for some initial states. However, it may be explained by the variations of all three parameters simultaneously with compensation phenomena between the right and left floodplains, which is commonly referred to as the double penalty effect [16]. Furthermore, the uniqueness of a solution of this type of problem is not mathematically proven (identifiability issues), and different sets of parameters can give analogous results.

V. CONCLUSIONS AND PERSPECTIVES

This work investigates the impact of dissimilarity measures on cost function evaluation in the context of automatic calibration using variational methods. The performance of these measures varied significantly, especially in calibrating channel versus floodplain friction parameters. This study underscores the importance of choosing the right dissimilarity measure to retrieve the appropriate parameter values, especially for the less sensitive parameters (e.g., the floodplains in this work). Among the tested measures, Normalized Mutual Information proved to be the most robust and efficient for calibration within the 3D-Var framework for this flood model, although it may require more iterations with some initial guesses.

The calibration results are affected by the choice of the dissimilarity measures, but also by the initial parameter

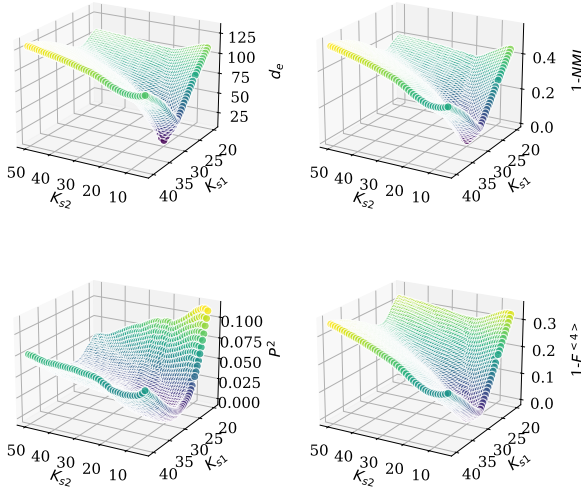


Fig. 9: Variation of the four dissimilarity measures as a function of the coupled parameters K_{s1} and K_{s2} for $K_{s3} = 10$.

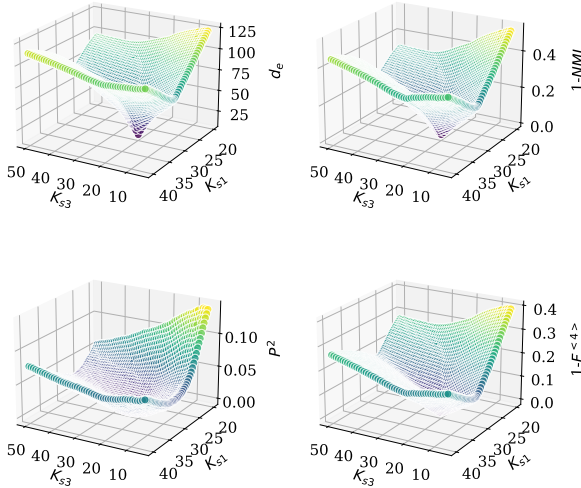


Fig. 10: Variation of the four dissimilarity measures as a function of the coupled parameters K_{s1} and K_{s3} for $K_{s2} = 20$.

guesses. Depending on the first initial guess of the parameter, some dissimilarity measures may or may not achieve the expected values during the calibration.

Future research should apply these methods to real-world flood events to confirm the influence of dissimilarity measures on flood model calibration. Another future research aims at including more noise in the data, since flood maps derived from satellite observations are often subject to uncertainties. The main hypothesis on the smoothness of the operator $\mathcal{H}$ should be investigated, and ensemble data assimilation methods may be more

efficient. Practitioners are encouraged to experiment with various dissimilarity measures, particularly for dense data with structures such as those from images and satellite observations, to optimize calibration processes.

VI. ACKNOWLEDGEMENT

This work was carried out as part of EDF R&D's MOISE-2 research project on river and coastal flood hazard assessment, whose support the authors gratefully acknowledge. This work was funded by the French National Association of Research and Technology (ANRT) and R&D through the Industrial Conventions for Training through REsearch (CIFRE agreement 2022/0972). The authors would also like to gratefully acknowledge the work of the open source communities of openTELEMAC and ADAO.

REFERENCES

- [1] Paul D Bates, Florian Pappenberger, and Renata J Romanowicz. Uncertainty in flood inundation modelling. In *Applied uncertainty analysis for flood risk management*, pages 232–269. 2014.
- [2] Eric Gilleland, David Ahijevych, Barbara G Brown, Barbara Casati, and Elizabeth E Ebert. Intercomparison of spatial forecast verification methods. *Weather and forecasting*, 24(5):1416–1430, 2009.
- [3] Stefania Grimaldi, Yuan Li, Valentijn Pauwels, and Jeffrey P Walker. Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges. *Surveys in Geophysics*, 37(5):977–1034, 2016.
- [4] Robert Manning, John Purser Griffith, TF Pigot, and Leveson Francis Vernon-Harcourt. *On the flow of water in open channels and pipes*. 1890.
- [5] Jean-Michel Hervouet. *Hydrodynamics of free surface flows: modelling with the finite element method*. John Wiley & Sons, 2007.
- [6] C Goeury, A Ponçot, JP Argaud, F Zaoui, R Ata, and Y Audouin. Optimal calibration of telemac-2d models based on a data assimilation algorithm. In *24th Telemac-Mascaret User Conference, Graz, Austria*, pages 73–80, 2017.
- [7] Renaud Hostache, Xijun Lai, Jerome Monnier, and Christian Puech. Assimilation of spatially distributed water levels into a shallow-water flood model. part ii: Use of a remote sensing image of mosel river. *Journal of hydrology*, 390(3-4):257–268, 2010.
- [8] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.
- [9] Hannah L Cloke and Florian Pappenberger. Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 15(1):181–197, 2008.
- [10] Jean-Paul Travert, Sébastien Boyaval, Cédric Goeury, Vito Bacchi, and Fabrice Zaoui. Evaluation of performance measures for qualifying flood models with satellite observations. *arXiv preprint arXiv:2408.00571*, 2024.
- [11] John C Gower. Generalized procrustes analysis. *Psychometrika*, 40:33–51, 1975.
- [12] Colin Studholme, Derek LG Hill, and David J Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition*, 32(1):71–86, 1999.
- [13] Jean-Paul Travert, Cédric Goeury, Vito Bacchi, Fabrice Zaoui, and Sébastien Boyaval. Comparing data assimilation algorithms for parameter estimation on telemac-2d.
- [14] Y Audouin, C Goeury, F Zaoui, R Ata, S El Idrissi Essebtay, A Torossian, and D Rouge. Interoperability applications of

- telemac-mascaret system. In *24th Telemac-Mascaret User Conference, Graz, Austria*, pages 57–64, 2017.
- [15] Jean-Philippe Argaud. User documentation, in the salome 9.3 platform, of the adao module for” data assimilation and optimization”. Technical report, Technical report 6125-1106-2019-01935-EN, EDF/R&D, 2019.
- [16] Alban Farchi, Marc Bocquet, Yelva Roustan, Anne Mathieu, and Arnaud Quérel. Using the wasserstein distance to compare fields of pollutants: application to the radionuclide atmospheric dispersion of the fukushima-daiichi accident. *Tellus B: Chemical and Physical Meteorology*, 68(1):31682, 2016.

A.2 Data assimilation with Optimal Transport on manifolds

A.2.1 Introduction

Classical implementations of variational data assimilation generally use $\mathcal{L}^2$ distances weighted by the inverse of error covariance matrices. Similarly, most Kalman-based data assimilation algorithms are formulated in the Euclidean space.

Although effective in addressing discrepancies in amplitude, the Euclidean space is often inadequate for capturing spatial misalignments. This can be problematic in applications where the structure and location of physical features, such as the position of a flood front, are critical. As noted by Hoffman et al. [1995], classical norms fail to fully describe such distortions.

Optimal transport provides a promising alternative through Wasserstein-type distances, which naturally quantify differences between distributions by accounting for both mass and displacement. In flood modeling, we hypothesize that using a Wasserstein distance in the data assimilation framework can lead to more accurate corrections of spatial errors and improve the alignment of water surface predictions with observations.

While data assimilation algorithms are usually derived on Euclidean spaces, we aim to assimilate probability distributions (or water depth distributions along a cross-section) on a manifold equipped with Wasserstein distance (Wasserstein spaces) [Panaretos and Zemel, 2020].

We illustrate the approach on a simplified one-dimensional shallow water model governed by the Saint-Venant equations. In this test case, the model is driven by a spatially varying friction coefficient, which we aim to infer using observations and prior knowledge through manifold-based data assimilation.

A.2.2 Data assimilation framework: Manifolds Extended Kalman Filter

The classical Extended Kalman Filter defined for Euclidean spaces can be extended to manifolds. The Manifolds Extended Kalman Filter (MEKF) is described more thoroughly in Le Ruz et al. [2025]. The algorithm writes starting from a state x_{0-} as:

Initialization

$$\mathbf{x}_{0-} \quad \text{and} \quad \Pi_{0-} = \Lambda_0^{-1}$$

Correction - Update

$$\begin{aligned} G_n &= \Pi_{n-} Dh(\hat{\mathbf{x}}_{n-})^T (\mathcal{R}^{-1} + \tau DH(\hat{\mathbf{x}}_{n-}) \Pi_{n-} DH(\hat{\mathbf{x}}_{n-})^T)^{-1}, \\ \hat{\mathbf{x}}_{n+} &= \exp_{\hat{\mathbf{x}}_{n-}} \left(\tau G_n \log_{H(\hat{\mathbf{x}}_{n-})} \mathbf{y}_n^{obs} \right), \\ \Pi_{n+} &= P_{\hat{\mathbf{x}}_{n-}, \hat{\mathbf{x}}_{n+}} ((Id - \tau G_n DH(\hat{\mathbf{x}}_{n-})) \Pi_{n-}) P_{\hat{\mathbf{x}}_{n-}, \hat{\mathbf{x}}_{n+}}^T. \end{aligned}$$

Prediction - Analysis

$$\begin{aligned} \hat{\mathbf{x}}_{n+1-} &= \mathcal{M}(\hat{\mathbf{x}}_{n+}), \\ \Pi_{n+1-} &= D\mathcal{M}(\hat{\mathbf{x}}_{n+}) \Pi_{n+} D\mathcal{M}(\hat{\mathbf{x}}_{n+})^T + \tau \mathcal{Q}. \end{aligned}$$

where:

- Π_{0-} is the initial covariance matrix on $\mathbf{x}_{0-}$

- $\mathcal{M}$ is the numerical model
- $D\mathcal{M}$ is the tangent of the model
- τ : Time step discretization
- $\mathcal{Q}$ model covariance error
- Π covariance error matrix on states
- H is the observation operator
- DH is the observation tangent operator
- $\mathcal{R}$ is the covariance observation error
- $\mathbf{y}_n^{obs}$ are the measurements at time n
- $\log$ and $\exp$ are the logarithm and exponential maps on the Riemannian manifold
- $P_{\hat{\mathbf{x}}_{n-}, \hat{\mathbf{x}}_{n+}}$ is the parallel transport between $\hat{\mathbf{x}}_{n-}$ and $\hat{\mathbf{x}}_{n+}$

Remark: To implement the algorithm to estimate the state and parameters of the numerical model on a manifold, the exponential and logarithm maps should be defined. The parallel transport between two vectors is computed based on the exponential and logarithm maps and is described in [Guigui and Pennec \[2022\]](#) for more details. In our case, the manifold is the space of probability measures (e.g., water depth distributions) equipped with the Wasserstein distance that will be described in Section [A.2.4](#).

A.2.3 Toy model

To validate the applicability of the Manifolds Extended Kalman Filter (MEKF) for calibrating flood models, we begin with a simplified one-dimensional (1D) test case. Although our ultimate goal is to address 2D flood model calibration, starting with a 1D model allows us to implement the method in a controlled and simpler setting.

We consider the Saint-Venant equations for shallow water flows, solved using a finite volume method with an HLLC (Harten–Lax–van Leer–Contact) flux solver. We also incorporate a bottom shear stress term modeled using the Manning–Strickler formulation to account for bed friction.

A.2.3.1 Governing equations

The 1D Saint-Venant equations are written as follows:

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} = 0, \\ \frac{\partial(hu)}{\partial t} + \frac{\partial(hu^2 + \frac{1}{2}gh^2)}{\partial x} = -gh \frac{\partial z_b}{\partial x} - g \frac{u|u|}{K_s^2 h^{4/3}}, \end{cases} \quad (\text{A.1})$$

where:

- $h(x, t)$ is the water depth,
- $u(x, t)$ is the depth-averaged velocity,

- $z_b(x)$ is the bed elevation,
- g is the gravitational acceleration,
- K_s is the Strickler roughness coefficient.

We can write in a more compact form:

$$\partial_t U + \partial_x F(U) = S(U), \quad (\text{A.2})$$

where $U = \begin{pmatrix} h \\ hu \end{pmatrix}$, $F(U) = \begin{pmatrix} hu \\ \frac{hu^2}{h} + \frac{gh^2}{2} \end{pmatrix}$ and $S(U) = \begin{pmatrix} 0 \\ -gh\partial_x z_b - g\frac{u|u|}{K_s^2 h^{4/3}} \end{pmatrix}$

A.2.3.2 Numerical scheme

The equations are discretized using a first-order finite volume approach. The update for cell averages is written as:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x} (\mathbf{F}_{i+1/2} - \mathbf{F}_{i-1/2}) + \Delta t \mathbf{S}_i, \quad (\text{A.3})$$

where $\mathbf{F}_{i\pm 1/2}$ is the HLLC flux computed at the interfaces.

HLLC Fluxes

The HLLC flux [Harten et al., 1983] at the interface $i + 1/2$ is computed based on left and right states $\mathbf{U}_L = \mathbf{U}_i$, $\mathbf{U}_R = \mathbf{U}_{i+1}$.

$$\mathbf{F}_{i+1/2} = \begin{cases} \mathbf{F}_L, & \text{if } 0 \leq c_1, \\ \frac{c_2 \mathbf{F}(\mathbf{U}_L) - c_1 \mathbf{F}(\mathbf{U}_R)}{c_2 - c_1} + \frac{c_1 c_2}{c_2 - c_1} (\mathbf{U}_R - \mathbf{U}_L) & \text{if } c_1 < 0 < c_2 \\ \mathbf{F}_R, & \text{if } c_2 \leq 0, \end{cases} \quad (\text{A.4})$$

with:

$$c_1 = \inf_{\mathbf{U}=\mathbf{U}_L, \mathbf{U}_D} \left(\inf_{j \in \{1;2\}} \lambda_j(\mathbf{U}) \right), \quad c_2 = \sup_{\mathbf{U}=\mathbf{U}_L, \mathbf{U}_D} \left(\sup_{j \in \{1;2\}} \lambda_j(\mathbf{U}) \right), \quad (\text{A.5})$$

where $\lambda_1(\mathbf{U})$ and $\lambda_2(\mathbf{U})$ are eigenvalues of the system.

A.2.3.3 Example of water depth distributions

In Figure A.1, we display an example of the observed water depths $h(x, t)$ for $t = 21600s$ for two parameterizations of the Strickler coefficient (uniform Strickler in the entire domain) and a given inlet discharge. The time step is fixed at 30 s, and the discretization is set to 500 m with a total of 40 nodes in the domain.

A.2.3.4 Application of MEKF

In this toy problem, the goal is to calibrate the Strickler coefficient $K_s(x)$ using the MEKF algorithm presented in Section A.2.2, by assimilating observations of water depths $h(x, t)$. We generate discrete measurements of h in time and space for a given parameterization of K_s , and the MEKF is used to estimate K_s along with state variables h and u . Then, in the algorithm defined in section A.2.2 the control vector $\mathbf{x}$ is as follows:

$$\hat{\mathbf{x}}_{n+} = [\hat{h}_{n+}^1, \dots, \hat{h}_{n+}^d, \hat{u}_{n+}^1, \dots, \hat{u}_{n+}^d, \hat{K}_{s_{n+}}^1, \dots, \hat{K}_{s_{n+}}^p]^T \in \mathbb{R}^N \quad (\text{A.6})$$

with d the number of nodes in the domain and p the number of friction zones.

As for the observations we have :

$$\mathbf{y}_n^{obs} = [(h^{obs})_n^1, \dots, (h^{obs})_n^d]^T \quad (\text{A.7})$$

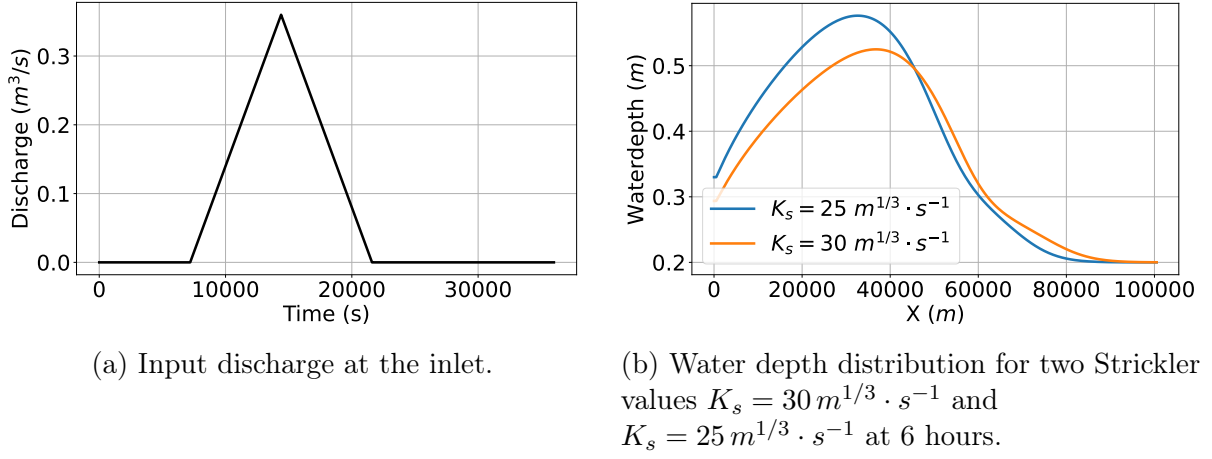


Figure A.1: Input discharge and water depths simulated by the 1D model for two Strickler configurations.

A.2.4 Optimal transport

To implement the data assimilation algorithm proposed by [Le Ruz et al. \[2025\]](#), the exponential and logarithm maps associated with the chosen Riemannian metric should be computed.

A.2.4.1 Monge problem

Gaspard Monge introduced optimal transport theory in 1781 [[Monge, 1781](#)]. His initial assignment was to move rubble in an embankment of the same volume as visualised in Figure A.2. The displacement of a mass has a cost, so the goal is to minimize the energy required to move rubble.

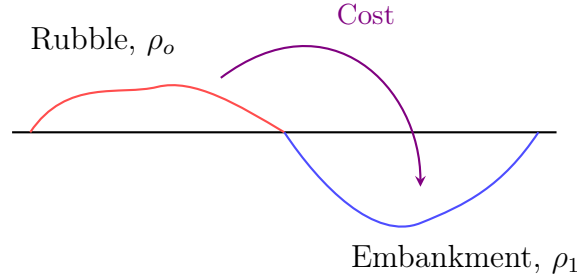


Figure A.2: Illustration of the Monge problem between two continuous distributions of same masses.

Let us define a quadratic cost c defined for any couple of points $(x, y) \in (\mathcal{X} \times \mathcal{Y})$:

$$c(x, y) = \|x - y\|_2^2. \quad (\text{A.8})$$

The movement of mass from the pile of rubble to the embankment can be described by a transport map $T : \mathcal{X} \rightarrow \mathcal{Y}$, and it induces a transport cost of:

$$C(T) = \int_{\mathcal{X}} c(x, T(x)) d\rho_0(x). \quad (\text{A.9})$$

The Monge problem is solved by finding the optimal transport map T (Monge map) and defines the Wasserstein distance:

$$\mathcal{W}^2 = \inf_{T: T\# \rho_0 = \rho_1} C(T), \quad (\text{A.10})$$

where $T\#\rho_0 = \rho_1$ is the push-forward of ρ_0 under T , or T pushes ρ_0 towards ρ_1 .

A.2.4.2 Monge problem between discrete measures

In geosciences, ρ_0 and ρ_1 are usually discrete measures (e.g., water depth distribution at the mesh nodes). Let us consider weights a and b with locations $x_1, \dots, x_n \in \mathcal{X}$, and $y_1, \dots, y_m \in \mathcal{Y}$ respectively then ρ_0 and ρ_1 reads:

$$\rho_0 = \sum_{i=1}^n a_i \delta_{x_i} \quad \rho_1 = \sum_{j=1}^m b_j \delta_{y_j} \quad (\text{A.11})$$

The Monge problem aims to push the mass ρ_0 to ρ_1 by associating each point x_i to a single point in y_j with a transport map $T : \{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_m\}$ verifying:

$$\forall j, b_j = \sum_{i: T(x_i)=y_j} a_i, \quad (\text{A.12})$$

usually written as $T\#\rho_0 = \rho_1$ with $T\#\rho_0 = \sum_i a_i \delta_{T(x_i)}$. Then the Wasserstein distance is defined as:

$$\min_T \left\{ \sum_i c(x_i, T(x_i)) : T\#\rho_0 = \rho_1 \right\} \quad (\text{A.13})$$

For discrete measures, the goal is then to find a permutation $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$ such that $j = \sigma(i)$ and $\sum_{i \in \sigma^{-1}(j)} a_i = b_j$. In Figure A.3, the optimal transport map T moving x_2 to x_{10} is visualized.

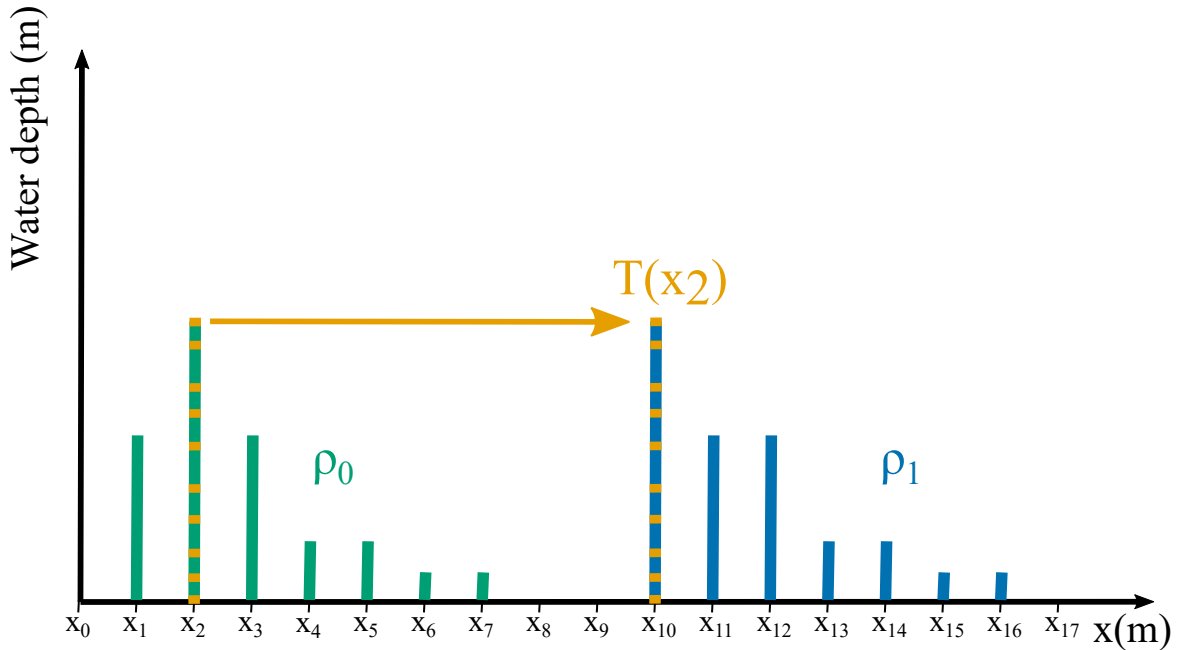


Figure A.3: Monge transport between two discrete distributions ρ_0 and ρ_1 .

For discrete measures, the weight vectors are not always compatible, and there may not exist such a transport map T .

A.2.4.3 Kantorovich relaxation

Monge maps may not exist between discrete measures with the same mass. The idea of Kantorovich is to relax the nature of the transportation by enabling mass splitting

[Kantorovich, 1942]. The mass at x_i can be dispatched across several locations instead of a single location. In place of a transport map T , a transport plan $\pi \in \mathbb{R}^{n \times m}$, where $\pi_{i,j}$ describes the amount of mass moving from position i toward position j . An example of mass splitting on water depth discrete measures is visualized in Figure A.4 The set of

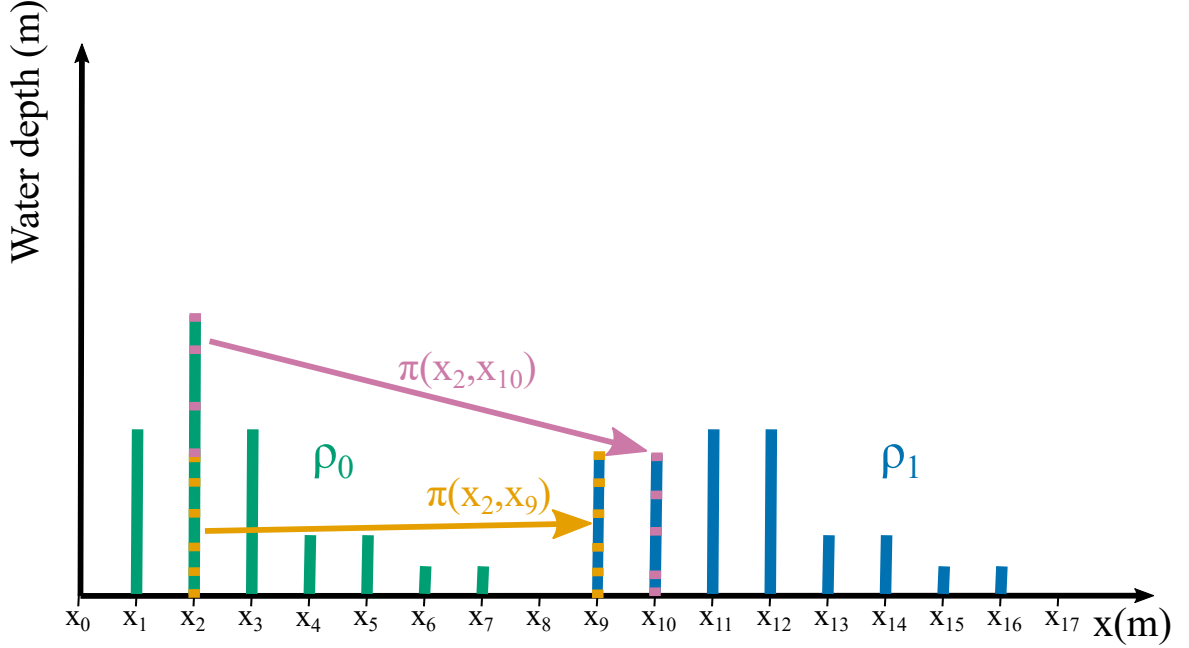


Figure A.4: Monge-Kantorovich transport plan with mass splitting between two discrete distributions ρ_0 and ρ_1 .

possible transport plans between ρ_0 and ρ_1 can be described as:

$$\pi(\rho_0, \rho_1) = \{\pi \in \mathbb{R}_+^{n \times m} : \pi \mathbb{1}_m = a \quad \text{and} \quad \pi^T \mathbb{1}_n = b\}. \quad (\text{A.14})$$

Kantorovich optimal transport problem then writes:

$$\mathcal{W}^2 = \min_{\pi \in \Pi(\rho_0, \rho_1)} C_{i,j} \pi_{i,j}, \quad (\text{A.15})$$

with $C_{i,j}$ the cost of moving from position x_i to y_j . This problem is usually solved using linear programming, but the optimal transport plan may not be unique. If $m = n$ and $\rho_0 = \frac{1}{n}(\delta(x_1), \delta(x_2), \dots, \delta(x_{n-1}), \delta(x_n))$ and $\rho_1 = \frac{1}{n}(\delta(y_1), \delta(y_2), \dots, \delta(y_{n-1}), \delta(y_n))$, then there exists an optimal solution π_{σ^*} to the Kantorovich problem, a permutation matrix associated with an optimal permutation σ^* solving the Monge problem.

Remark: We aim to compare water depth probability distributions. The weights of the simulated and observed probability distributions may not be compatible with solving the Monge problem; therefore, we have chosen to use the Monge-Kantorovich formulation with mass splitting.

A.2.4.4 Geodesics, Logarithm Map and Exponential Map in Wasserstein Spaces

The space of probability measures equipped with the Wasserstein distance defined via the Monge-Kantorovich problem forms a metric space known as the Wasserstein space. We apply the MEKF on this manifold denoted as $\mathcal{W}_2(\mathbb{R})$. The exponential and logarithm maps are defined for continuous measures on this space.

Continuous measures Let us consider $\rho_0 \in \mathcal{W}_2(\mathbb{R}^d)$ and $\rho_1 \in \mathcal{W}_2(\mathbb{R}^d)$ two continuous measures. We consider that a Monge map T exists between the two distributions. We can define a curve known as McCann’s interpolant, which defines a geodesic with displacement interpolation between the two distributions [McCann, 1997]:

$$(\rho_0)_t = [Id + t(T - Id)]\#\rho_0 \quad t \in [0, 1]. \quad (\text{A.16})$$

For this interpolant, we have $\rho_0 = (\rho_0)_0$ and $\rho_1 = (\rho_0)_1$. This path defines the geodesic on the Wasserstein space.

Logarithm map The logarithm map enables us to determine the tangent vector at the initial point, allowing us to transition from one distribution to another. It is defined as:

$$\log_{\rho_0}(\rho_1) = T - Id \quad (\text{A.17})$$

Exponential map Let us consider v a velocity field from the tangent space at ρ_0 , then:

$$\exp_{\rho_0}(v) = (Id + v)\#\rho_0 \quad (\text{A.18})$$

For continuous measures, McCann’s interpolant can be extended with a transport plan π , but the logarithm map may not be defined (to our knowledge).

Discrete measures Let us consider now that ρ_0 and ρ_1 are discrete probability measures. We should note that the space of discrete probability measures is not a smooth manifold. There’s usually no Monge map, only a transport plan π .

Remark: The tangent space seems ill-defined, and we cannot construct a logarithm map as for the continuous measures.

Discrete logarithm map In our implementation, we choose the transport plan π to play the role of the logarithm, encoding the entire direction of movement from ρ_0 to ρ_1 ; however, it is not the "true" logarithm map.

Discrete geodesics From the transport vectors v_i :

$$(\rho_0)_t = \sum_{i=1}^n a_i \delta_{x_i + tv_i}. \quad (\text{A.19})$$

Note: We encounter difficulty in implementing the MEKF because the logarithm map on the space of discrete probability measures equipped with the Wasserstein distance (in Monge-Kantorovich) is not clearly defined.

A.2.5 Computation of the exponential, logarithm, and parallel transport on the application case

A.2.5.1 Implementation in the code

We did not find a proper definition of the logarithm map for discrete measures, but we have attempted to replicate its behavior using the transport plan π in place of an accurate logarithm map.

Exponential Map Given a base distribution $\rho_0 \in \mathbb{R}^n$ and a “velocity field” $\pi \in \mathbb{R}^{n \times n}$ representing a transport plan, the exponential map $\exp_{\rho_0}(t\pi)$ moves the mass from ρ_0 along the “direction” π , scaled by a parameter $t \in [0, 1]$. Let the base distribution be

denoted by $\rho_0 = ((\rho_0)_1, \dots, (\rho_0)_n)$, and π_{ij} denote the mass transported from location x_i to location x_j . The pushforward location for each transported element is:

$$x_{ij}^{(t)} = (1 - t)i + tj. \quad (\text{A.20})$$

Since this position is not necessarily an integer, linear interpolation is used to move the mass onto the nearest point on the model discretization:

$$\begin{aligned} x_{\text{int}} &= \lfloor x_{ij}^{(t)} \rfloor \\ \alpha &= x_{ij}^{(t)} - x_{\text{int}} \end{aligned} \quad (\text{A.21})$$

Then, the resulting distribution $(\rho_0)_t$ after applying the exponential map is updated as:

$$(\rho_0)_t[x_{\text{int}}] += (1 - \alpha)v_{ij}, \quad (\rho_0)_t[x_{\text{int}} + 1] += \alpha v_{ij}. \quad (\text{A.22})$$

Logarithmic Map Given a source distribution ρ_0 and a target distribution ρ_1 , the logarithmic map computes a velocity field v such that:

$$\rho_1 = \exp_{\rho_0}(v). \quad (\text{A.23})$$

To compute v , we solve the optimal transport problem:

$$v = \pi_{i,j} - Id, \quad (\text{A.24})$$

where π is the transport plan (solved here using Python Optimal Toolbox [Flamary et al., 2021]). The diagonal subtraction ensures the mass remaining in place is accounted for, leading to a transport plan representing the velocity from ρ_0 to ρ_1 .

A.2.5.2 Analysis of the results

We computed the exponential and logarithm maps on the water depth distributions at the different times of observation based on Sections A.2.4.4 and A.2.5.1. Figure A.5 shows an example of geodesic and transport map for one of the observation time using two Strickler values in the domain ($K_s = 25m^{1/3} \cdot s^{-1}$ and $K_s = 30m^{1/3} \cdot s^{-1}$) for generating the source and target distributions (observed and simulated water depth distributions).

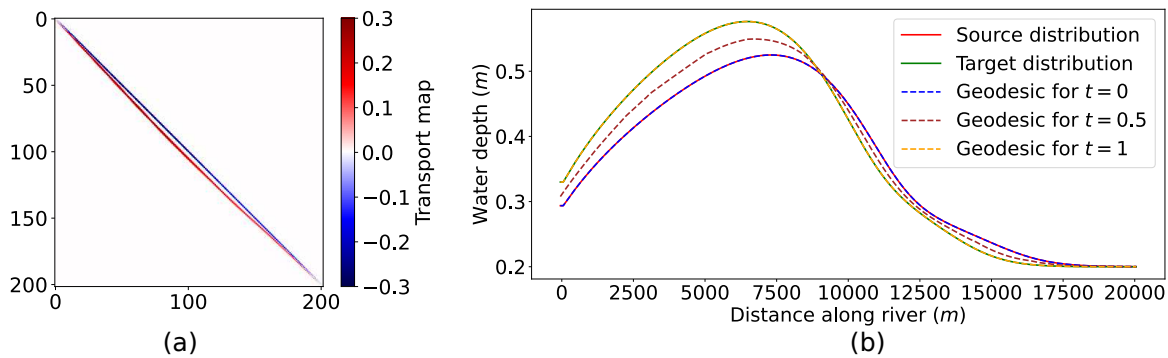


Figure A.5: (a) Transport map and (b) geodesics between two water depth distributions computed with a one-dimensional shallow-water code.

Then, we implemented the parallel transport based on the algorithm in Guigui and Pennec [2022]. Parallel transport along a geodesic curve transports a vector from the tangent space at a point to another tangent space at a second point. To test the parallel transport computation, we computed the logarithm map between a simulation ran with

target distribution using $K_s = 30 m^{1/3} \cdot s^{-1}$ and source distribution $K_s = 20 m^{1/3} \cdot s^{-1}$. Using the transport map between the two distributions we compute the parallel transport between a distribution with $K_s = 20 m^{1/3} \cdot s^{-1}$ and a third simulation with $K_s = 25 m^{1/3} \cdot s^{-1}$. Using the exponential map with the transported vector resulting from the parallel transport with the distribution generated with $K_s = 25 m^{1/3} \cdot s^{-1}$ should result in a distribution which coincide with the target distribution generated with $K_s = 30 m^{1/3} \cdot s^{-1}$. Figure A.6 shows the three distributions and the distribution that used the transport vector.

The distribution computed with the exponential map from the water depth distribution generated with $K_s = 25 m^{1/3} \cdot s^{-1}$ does not coincide with the water target distribution generated with $K_s = 30 m^{1/3} \cdot s^{-1}$. The application of parallel transport was unsuccessful, as the logarithm map for discrete measures using the Monge-Kantorovich formulation was ill-defined. The logarithm map used in the implementation does not live in the tangent space of the Wassertein space. Indeed, when taking two transport plans π_1 and π_2 , the transport plan $\pi = \pi_1 + \pi_2$ does not result in a transport plan (the mass of the distribution is not conserved).

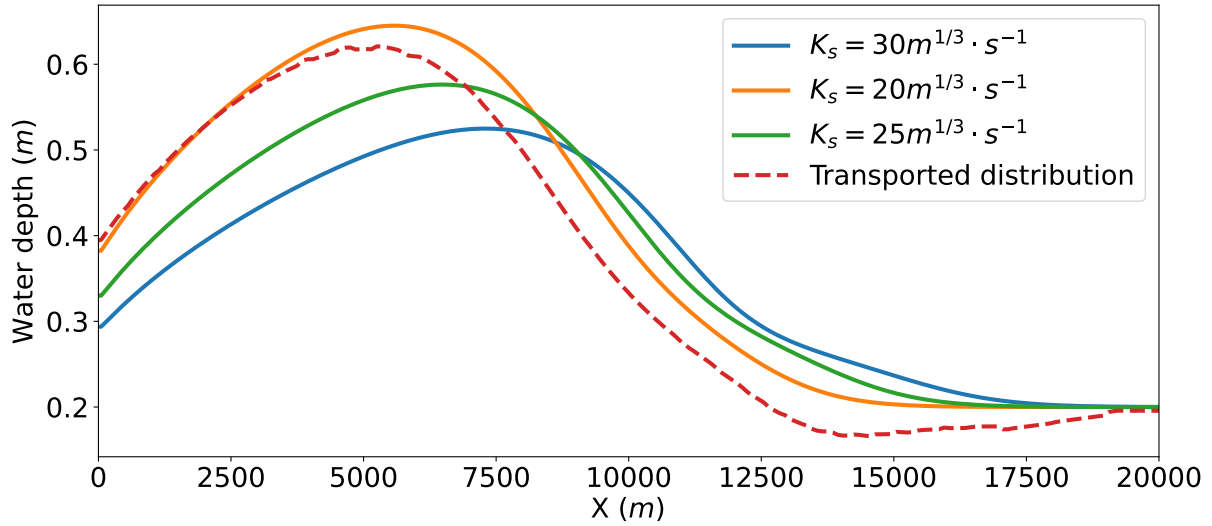


Figure A.6: Verification of parallel transport implementation.

Bibliography

- AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*, 53(2):452–480. 37
- Alsdorf, D., Bates, P., Melack, J., Wilson, M., and Dunne, T. (2007a). Spatial and temporal complexity of the amazon flood measured from space. *Geophysical research letters*, 34(8). 57
- Alsdorf, D. E., Rodríguez, E., and Lettenmaier, D. P. (2007b). Measuring surface water from space. *Reviews of Geophysics*, 45(2). 40
- Alsdorf, D. E., Smith, L. C., and Melack, J. M. (2001). Amazon floodplain water level changes measured with interferometric sir-c radar. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):423–431. 16, 57
- Arcement, G. J. and Schneider, V. R. (1989). *Guide for selecting Manning’s roughness coefficients for natural channels and flood plains*. US Government Printing Office Washington, DC. 15, 30
- Aronica, G., Bates, P., and Horritt, M. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within glue. *Hydrological processes*, 16(10):2001–2016. 36
- Asch, M., Bocquet, M., and Nodet, M. (2016). *Data assimilation: methods, algorithms, and applications*. SIAM. 74, 77, 79, 91
- Autorité de Sûreté Nucléaire (2013). Protection of basic nuclear installations against external flooding. Technical report, Autorité de Sûreté Nucléaire. 17
- Ayvaz, M. T. (2013). A linked simulation–optimization model for simultaneously estimating the manning’s surface roughness values and their parameter structures in shallow water flows. *Journal of hydrology*, 500:183–199. 36
- Bacchi, S. and Pavan, S. (2025). Guide méthodologique pour les études sri-cgb / ror : Construction de modèles et déroulé d’une étude inondation en hydraulique fluviale. Technical report, EDF R&D. 13, 28, 30, 35
- Balbi, M. and Lallemand, D. C. B. (2023). Bayesian calibration of a flood simulator using binary flood extent observations. *Hydrology and Earth System Sciences*, 27(5):1089–1108. 37
- Ballesteros, J., Bodoque, J., Díez-Herrero, A., Sanchez-Silva, M., and Stoffel, M. (2011). Calibration of floodplain roughness and estimation of flood discharge based on tree-ring evidence and hydraulic modelling. *Journal of Hydrology*, 403(1-2):103–115. 36
- Ban, Y. and Yousif, O. A. (2012). Multitemporal spaceborne sar data for urban change detection in china. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4):1087–1094. 53

- Barnes, H. H. (1967). *Roughness characteristics of natural channels*. Number 1849. US Government Printing Office. [15](#), [29](#), [30](#), [221](#)
- Bates, P. D., Pappenberger, F., and Romanowicz, R. J. (2014). Uncertainty in flood inundation modelling. In *Applied uncertainty analysis for flood risk management*, pages 232–269. [15](#)
- Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A.-L. (2015). Open turns: An industrial software for uncertainty quantification in simulation. *arXiv preprint arXiv:1501.05242*. [10](#), [67](#)
- Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55. [37](#)
- Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L., et al. (2018). Toward global soil moisture monitoring with sentinel-1: Harnessing assets and overcoming obstacles. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):520–539. [56](#)
- Bayes, T. (1763). Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. *Philosophical transactions of the Royal Society of London*, pages 370–418. [78](#)
- Bazi, Y., Bruzzone, L., and Melgani, F. (2005). An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):874–887. [53](#)
- Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16):4345–4378. [54](#)
- Bessis, J.-L., Bequignon, J., and Mahmood, A. (2004). The international charter “space and major disasters” initiative. *Acta Astronautica*, 54(3):183–190. [16](#)
- Betterle, A. and Salamon, P. (2024). Water depth estimate and flood extent enhancement for satellite-based inundation maps. *Natural Hazards and Earth System Sciences*, 24(8):2817–2836. [58](#), [218](#)
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology*, 320(1-2):18–36. [36](#)
- Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M. (2016). The swot mission and its capabilities for land hydrology. In *Remote sensing and water resources*, pages 117–147. Springer. [231](#)
- Blatman, G. and Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics*, 230(6):2345–2367. [72](#)
- Blayo, E., Cosme, E., Nodet, M., and Vidart, A. (2011). Introduction to data assimilation. Technical report, Tech. rep. [61](#), [62](#)
- Bodart, G. (2023). *Evaluation et amélioration des méthodes de détermination des vitesses en surface d’écoulement par analyse de séquence d’images*. PhD thesis, Université Grenoble Alpes [2020-....]. [11](#), [232](#)
- Bonafilia, D., Tellman, B., Anderson, T., and Issenberg, E. (2020). Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 210–211. [54](#)

- Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784. 70, 73
- Bovolo, F. and Bruzzone, L. (2007). A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1658–1670. 53
- Brakenridge, G., Tracy, B., and Knox, J. (1998). Orbital sar remote sensing of a river flood wave. *International journal of remote sensing*, 19(7):1439–1445. 53, 57, 59
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32. 54
- Brown, K. M., Hambidge, C. H., and Brownett, J. M. (2016). Progress in operational flood mapping using satellite synthetic aperture radar (sar) and airborne light detection and ranging (lidar) data. *Progress in Physical Geography*, 40(2):196–214. 58, 59
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90. 91
- Bryant, S., McGrath, H., and Boudreault, M. (2022). Gridded flood depth estimates from satellite-derived inundations. *Natural Hazards and Earth System Sciences*, 22(4):1437–1450. 59
- Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724. 83
- Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208. 91, 220
- Cacuci, D. G. (2003). *Sensitivity & Uncertainty Analysis, Volume 1: Theory (1st ed.)*. Chapman and Hall/CRC. 69
- Calmant, S., Seyler, F., and Cretaux, J. F. (2008). Monitoring continental surface waters by satellite altimetry. *Surveys in geophysics*, 29(4):247–269. 40
- Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10):1509–1518. 70
- Cao, L., Coops, N. C., Sun, Y., Ruan, H., Wang, G., Dai, J., and She, G. (2019). Estimating canopy structure and biomass in bamboo forests using airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148:114–129. 37
- Cao, Y., Zhu, J., Navon, I. M., and Luo, Z. (2007). A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53(10):1571–1583. 89
- Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535. 61, 62, 74, 81, 82, 83
- Cazals, C., Rapinel, S., Frison, P.-L., Bonis, A., Mercier, G., Mallet, C., Corgne, S., and Rudant, J.-P. (2016). Mapping and characterization of hydrological dynamics in a coastal marsh using high temporal resolution sentinel-1a images. *Remote Sensing*, 8(7):570. 55
- Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277. 53

- Cheng, S., Prentice, I. C., Huang, Y., Jin, Y., Guo, Y.-K., and Arcucci, R. (2022). Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. *Journal of Computational Physics*, 464:111302. 74
- Chini, M., Hostache, R., Giustarini, L., and Matgen, P. (2017). A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):6975–6988. 53
- Chini, M., Pelich, R., Pulvirenti, L., Pierdicca, N., Hostache, R., and Matgen, P. (2019). Sentinel-1 insar coherence to detect floodwater in urban areas: Houston and hurricane harvey as a test case. *Remote sensing*, 11(2):107. 55
- Chow, V. (1959). *Open Channel Hydraulics*. McGraw-Hill, New York. 15, 23, 29, 30, 31, 32, 35, 221
- Chuvieco, E. (2020). *Fundamentals of satellite remote sensing: An environmental approach*. CRC press. 37
- Cian, F., Marconcini, M., and Ceccato, P. (2018a). Normalized difference flood index for rapid flood mapping: Taking advantage of eo big data. *Remote Sensing of Environment*, 209:712–730. 55
- Cian, F., Marconcini, M., Ceccato, P., and Giupponi, C. (2018b). Flood depth estimation by means of high-resolution sar images and lidar data. *Natural Hazards and Earth System Sciences*, 18(11):3063–3084. 58, 59
- Clement, M. A., Kilsby, C., and Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, 11(2):152–168. 54, 56
- Cohen, S., Brakenridge, G. R., Kettner, A., Bates, B., Nelson, J., McDonald, R., Huang, Y.-F., Munasinghe, D., and Zhang, J. (2018). Estimating floodwater depths from flood inundation maps and topography. *JAWRA Journal of the American Water Resources Association*, 54(4):847–858. 57, 59
- Cohen, S., Raney, A., Munasinghe, D., Loftis, J. D., Molthan, A., Bell, J., Rogers, L., Galantowicz, J., Brakenridge, G. R., Kettner, A. J., et al. (2019). The floodwater depth estimation tool (fwdet v2. 0) for improved remote sensing analysis of coastal flooding. *Natural Hazards and Earth System Sciences*, 19(9):2053–2065. 57, 59
- Cooper, E. S., Dance, S. L., García-Pintado, J., Nichols, N. K., and Smith, P. J. (2019). Observation operators for assimilation of satellite observations in fluvial inundation forecasting. *Hydrology and Earth System Sciences*, 23(6):2541–2559. 16, 232
- Copernicus (2022). Copernicus emergency management service. 16
- Copernicus Emergency Management Service (2015). Emsr149: Flood in ireland - activation overview. <https://emergency.copernicus.eu/mapping/list-of-components/EMSR149>. Accessed: 2025-07-07. 126
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A. (1994). A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387. 88
- Cowan, W. L. (1956). Estimating hydraulic roughness coefficients. *Agricultural engineering*, 37(7):473–475. 32
- Crétaux, J.-F. and Birkett, C. (2006). Lake studies from satellite radar altimetry. *Comptes Rendus Geoscience*, 338(14-15):1098–1112. 40

- Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., and Arheimer, B. (2020). Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, 65(5):699–711. 15
- Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021). *Basics and trends in sensitivity analysis: Theory and practice in R*. SIAM. 69
- D’Addabbo, A., Refice, A., Pasquariello, G., Lovergine, F. P., Capolongo, D., and Manfreda, S. (2016). A bayesian network for flood detection combining sar imagery and ancillary data. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3612–3625. 54
- Dalledonne, G. L., Kopmann, R., and Brudy-Zippelius, T. (2019). Uncertainty quantification of floodplain friction in hydrodynamic models. *Hydrology and Earth System Sciences*, 23(8):3373–3385. 30
- Dalsasso, E., Denis, L., and Tupin, F. (2021a). As if by magic: Self-supervised training of deep despeckling networks with merlin. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13. 46
- Dalsasso, E., Denis, L., and Tupin, F. (2021b). Sar2sar: A semi-supervised despeckling algorithm for sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4321–4329. 46
- Dasgupta, A., Grimaldi, S., Ramsankaran, R., Pauwels, V. R., and Walker, J. P. (2018). Towards operational sar-based flood mapping using neuro-fuzzy texture-based approaches. *Remote sensing of Environment*, 215:313–329. 54
- Dasgupta, A., Grimaldi, S., Ramsankaran, R., Pauwels, V. R., and Walker, J. P. (2022). A simple framework for calibrating hydraulic flood inundation models using crowd-sourced water levels. *Journal of Hydrology*, 614:128467. 231
- Dasgupta, A., Hostache, R., Ramsankaran, R., Grimaldi, S., Matgen, P., Chini, M., Pauwels, V. R., and Walker, J. P. (2021a). Earth observation and hydraulic data assimilation for improved flood inundation forecasting. In *Earth Observation for Flood Applications*, pages 255–294. Elsevier. 36
- Dasgupta, A., Hostache, R., Ramsankaran, R., Schumann, G. J.-P., Grimaldi, S., Pauwels, V. R., and Walker, J. P. (2021b). A mutual information-based likelihood function for particle filter flood extent assimilation. *Water Resources Research*, 57(2):e2020WR027859. 16, 233
- De Doncker, L., Troch, P., Verhoeven, R., Bal, K., Meire, P., and Quintelier, J. (2009). Determination of the manning roughness coefficient influenced by vegetation in the river aa and bieber river. *Environmental fluid mechanics*, 9:549–567. 33
- de Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. John Wiley & Sons. 66
- de Saint-Venant, A. J.-C. (1871). Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à l’introduction des marées dans leur lit. *CR Acad. Sci. Paris*, 73(147-154):237–240. 26
- Deems, J. S., Painter, T. H., and Finnegan, D. C. (2013). Lidar measurement of snow depth: a review. *Journal of Glaciology*, 59(215):467–479. 37
- Deledalle, C.-A., Denis, L., Tabti, S., and Tupin, F. (2017). Mulog, or how to apply gaussian denoisers to multi-channel sar speckle reduction? *IEEE Transactions on Image Processing*, 26(9):4389–4403. 46

- Deledalle, C.-A., Denis, L., Tupin, F., Reigber, A., and Jäger, M. (2014). Nl-sar: A unified nonlocal framework for resolution-preserving (pol)(in) sar denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2021–2038. 46
- Di Baldassarre, G., Schumann, G., Brandimarte, L., and Bates, P. (2011). Timely low resolution sar imagery to support floodplain modelling: a case study review. *Surveys in geophysics*, 32(3):255–269. 54
- Di Martino, G., Di Simone, A., Iodice, A., and Riccio, D. (2016). Scattering-based nonlocal means sar despeckling. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3574–3588. 46
- Direction de la Climatologie, M. (2009). *Statistiques climatiques de la France 1971-2000*. Météo-France, Direction de la climatologie. 40
- Domeneghetti, A., Castellarin, A., Tarpanelli, A., and Moramarco, T. (2015). Investigating the uncertainty of satellite altimetry products for hydrodynamic modelling. *Hydrological Processes*, 29(23):4908–4918. 16
- Domeneghetti, A., Molari, G., Tourian, M. J., Tarpanelli, A., Behnia, S., Moramarco, T., Sneeuw, N., and Brath, A. (2021). Testing the use of single-and multi-mission satellite altimetry for the calibration of hydraulic models. *Advances in Water Resources*, 151:103887. 42
- Domeneghetti, A., Tarpanelli, A., Brocca, L., Barbetta, S., Moramarco, T., Castellarin, A., and Brath, A. (2014). The use of remote sensing-derived water surface data for hydraulic model calibration. *Remote sensing of environment*, 149:130–141. 16, 36, 42
- Edwards, C. A., Moore, A. M., Hoteit, I., and Cornuelle, B. D. (2015). Regional ocean data assimilation. *Annual review of marine science*, 7(1):21–42. 74
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, 32(2):407–499. With discussion, and a rejoinder by the authors. 72
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162. 81, 83
- Evensen, G. (2009a). *Data assimilation: the ensemble Kalman filter*. Springer. 81
- Evensen, G. (2009b). The ensemble kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29(3):83–104. 82
- Faivre, R., Makowski, D., Mahévas, S., and Iooss, B. (2013). Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l’environnement. 69
- Fewtrell, T., Bates, P. D., Horritt, M., and Hunter, N. (2008). Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes: An International Journal*, 22(26):5107–5118. 14
- Feyoux, N. (2016). *Transport optimal pour l’assimilation de données images*. PhD thesis, Université Grenoble Alpes. 61, 62
- Fix, E. and Hodges, J. (1951). Discriminatory analysis, nonparametric discrimination. 54
- Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., and Vayer, T. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8. 254

- Fletcher, R. (1970). A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322. 91
- Flores-Anderson, A. I., Herndon, K. E., Thapa, R. B., and Cherrington, E. (2019). The sar handbook: comprehensive methodologies for forest monitoring and biomass estimation. Technical report, Incomplet. 13, 49
- Forzieri, G., Degetto, M., Righetti, M., Castelli, F., and Preti, F. (2011). Satellite multispectral data for improved floodplain roughness modelling. *Journal of hydrology*, 407(1-4):41–57. 34
- French (1985). Open-channel. *Open-channel hydraulics*. 30
- Frost, V. S., Stiles, J. A., Shanmugan, K. S., and Holtzman, J. C. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on pattern analysis and machine intelligence*, (2):157–166. 46
- Fujita, I., Muste, M., and Kruger, A. (1998). Large-scale particle image velocimetry for flow analysis in hydraulic engineering applications. *Journal of hydraulic Research*, 36(3):397–414. 231
- Gagnon, L. and Jouan, A. (1997). Speckle filtering of sar images: a comparative study between complex-wavelet-based and standard filters. In *Wavelet Applications in Signal and Image Processing V*, volume 3169, pages 80–91. SPIE. 46
- Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014). Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603. 72
- Gat, I., Remez, T., Shaul, N., Kreuk, F., Chen, R. T., Synnaeve, G., Adi, Y., and Lipman, Y. (2024). Discrete flow matching. *Advances in Neural Information Processing Systems*, 37:133345–133385. 234
- Gerbeau, J.-F. and Perthame, B. (2000). *Derivation of viscous Saint-Venant system for laminar shallow water; numerical validation*. PhD thesis, INRIA. 23, 26
- Ghanem, R. G. and Spanos, P. D. (2003). *Stochastic finite elements: a spectral approach*. Courier Corporation. 72
- Ghil, M., Cohn, S., Tavantzis, J., Bube, K., and Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. *Dynamic meteorology: Data assimilation methods*, pages 139–224. 81
- Ghil, M. and Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. In *Advances in geophysics*, volume 33, pages 141–266. Elsevier. 81
- Giering, R. and Kaminski, T. (1998). Recipes for adjoint code construction. *ACM Transactions on Mathematical Software (TOMS)*, 24(4):437–474. 92
- Giglio, L., Schroeder, W., and Justice, C. O. (2016). The collection 6 modis active fire detection algorithm and fire products. *Remote sensing of environment*, 178:31–41. 37
- Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E. (2009). Intercomparison of spatial forecast verification methods. *Weather and forecasting*, 24(5):1416–1430. 228
- Giustarini, L., Hostache, R., Kavetski, D., Chini, M., Corato, G., Schlaffer, S., and Matgen, P. (2016). Probabilistic flood mapping using synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):6958–6969. 54

- Giustarini, L., Hostache, R., Matgen, P., Schumann, G. J.-P., Bates, P. D., and Mason, D. C. (2012). A change detection approach to flood mapping in urban areas using terrasars-x. *IEEE transactions on Geoscience and Remote Sensing*, 51(4):2417–2430. 53, 55
- Gobeyn, S., Van Wesemael, A., Neal, J., Lievens, H., Van Eerdenbrugh, K., De Vleeschouwer, N., Vernieuwe, H., Schumann, G. J.-P., Di Baldassarre, G., De Baets, B., et al. (2017). Impact of the timing of a sar image acquisition on the calibration of a flood inundation model. *Advances in Water Resources*, 100:126–138. 37
- Goeury, C., Audouin, Y., and Zaoui, F. (2022). Interoperability and computational framework for simulating open channel hydraulics: Application to sensitivity analysis and calibration of gironde estuary model. *Environmental Modelling & Software*, 148:105243. 9, 28
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26. 91
- Goodman, J. W. (1976). Some fundamental properties of speckle. *JOSA*, 66(11):1145–1150. 46
- Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE proceedings F (radar and signal processing)*, volume 140, pages 107–113. IET. 84
- Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005). Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer. 73
- Griewank, A. and Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. SIAM. 91
- Grimaldi, S., Li, Y., Pauwels, V., and Walker, J. P. (2016). Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges. *Surveys in Geophysics*, 37(5):977–1034. 16, 39, 41, 74
- Grimaldi, S., Xu, J., Li, Y., Pauwels, V. R., and Walker, J. P. (2020). Flood mapping under vegetation using single sar acquisitions. *Remote sensing of Environment*, 237:111582. 40, 55, 56
- Guigui, N. and Pennec, X. (2022). Numerical accuracy of ladder schemes for parallel transport on manifolds. *Foundations of Computational Mathematics*, 22(3):757–790. 248, 254
- Hall, J. W., Manning, L. J., and Hankin, R. K. (2011). Bayesian calibration of a flood inundation model using spatial data. *Water Resources Research*, 47(5). 37
- Harten, A., Lax, P. D., and Leer, B. v. (1983). On upstream differencing and godunov-type schemes for hyperbolic conservation laws. *SIAM review*, 25(1):35–61. 249
- Hascoet, L. and Pascual, V. (2013). The tapenade automatic differentiation tool: Principles, model, and specification. *ACM Transactions on Mathematical Software (TOMS)*, 39(3):1–43. 92
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. 78
- Henry, J.-B., Chastanet, P., Fellah, K., and Desnos, Y.-L. (2006). Envisat multi-polarized asar data for flood mapping. *International Journal of Remote Sensing*, 27(10):1921–1929. 51

- Heremans, R., Willekens, A., Borghys, D., Verbeeck, B., Valckenborgh, J., Acheroy, M., and Perneel, C. (2003). Automatic detection of flooded areas on envisat/asar images using an object-oriented classification technique and an active contour algorithm. In *International Conference on Recent Advances in Space Technologies, 2003. RAST'03. Proceedings of*, pages 311–316. IEEE. 53
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049. 74
- Hervouet, J.-M. (2007). *Hydrodynamics of free surface flows: modelling with the finite element method*. John Wiley & Sons. 27
- Hervouet, J.-M., Pavan, S., and Ricchiuto, M. (2017). *Residual distribution advection schemes in Telemac*. PhD thesis, Inria Bordeaux Sud-Ouest. 28
- Hoeffding, W. (1992). A class of statistics with asymptotically normal distribution. *Breakthroughs in statistics: Foundations and basic theory*, pages 308–334. 71
- Hoffman, R. N., Liu, Z., Louis, J.-F., and Grassoti, C. (1995). Distortion representation of forecast errors. *Monthly Weather Review*, 123(9):2758–2770. 247
- Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17. 71
- Horritt, M. (1999). A statistical active contour model for sar image segmentation. *Image and Vision Computing*, 17(3-4):213–224. 53
- Horritt, M. and Bates, P. (2001). Predicting floodplain inundation: raster-based modelling versus the finite-element approach. *Hydrological processes*, 15(5):825–842. 36
- Horritt, M. and Bates, P. (2002). Evaluation of 1d and 2d numerical models for predicting river flood inundation. *Journal of hydrology*, 268(1-4):87–99. 35, 36
- Horritt, M., Mason, D., and Luckman, A. (2001). Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *International Journal of Remote Sensing*, 22(13):2489–2507. 53
- Horritt, M. S. (2000). Calibration of a two-dimensional finite element flood flow model using satellite radar imagery. *Water Resources Research*, 36(11):3279–3291. 37
- Horton, R. E. (1916). Some better kutler's formula coefficients. *Eng. News Rec*, 75:373–374. 31
- Hossain, A. A., Jia, Y., and Chao, X. (2009). Estimation of manning's roughness coefficient distribution for hydrodynamic model using remotely sensed land cover features. In *2009 17th International Conference on Geoinformatics*, pages 1–4. IEEE. 34
- Hostache, R., Chini, M., Giustarini, L., Neal, J., Kavetski, D., Wood, M., Corato, G., Pelich, R.-M., and Matgen, P. (2018). Near-real-time assimilation of sar-derived flood maps for improving flood forecasts. *Water Resources Research*, 54(8):5516–5535. 16, 233
- Hostache, R., Lai, X., Monnier, J., and Puech, C. (2010). Assimilation of spatially distributed water levels into a shallow-water flood model. part ii: Use of a remote sensing image of mosel river. *Journal of hydrology*, 390(3-4):257–268. 36
- Hostache, R., Matgen, P., Schumann, G., Puech, C., Hoffmann, L., and Pfister, L. (2009). Water level estimation and reduction of hydraulic model calibration uncertainties using satellite sar images of floods. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):431–441. 16, 53, 58, 59, 228

- Hostache, R., Matgen, P., and Wagner, W. (2012). Change detection approaches for flood extent mapping: How to select the most adequate reference image from online archives? *International journal of applied earth observation and geoinformation*, 19:205–213. 53
- Huang, C., Nguyen, B. D., Zhang, S., Cao, S., and Wagner, W. (2017). A comparison of terrain indices toward their ability in assisting surface water mapping from sentinel-1 data. *ISPRS International Journal of Geo-Information*, 6(5):140. 56
- Huang, T., Yang, G., and Tang, G. (1979). A fast two-dimensional median filtering algorithm. *IEEE transactions on acoustics, speech, and signal processing*, 27(1):13–18. 46
- Huang, Y. and Qin, X. (2014). Uncertainty analysis for flood inundation modelling with a random floodplain roughness field. *Environmental Systems Research*, 3(1):1–7. 34
- Iooss, B. and Lemaître, P. (2015). A review on global sensitivity analysis methods. *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, pages 101–122. 69
- Jafarzadegan, K., Abbaszadeh, P., and Moradkhani, H. (2021). Sequential data assimilation for real-time probabilistic flood inundation mapping. *Hydrology and Earth System Sciences Discussions*, 2021:1–39. 83
- Jarrett, R. D. (1984). Hydraulics of high-gradient streams. *Journal of hydraulic engineering*, 110(11):1519–1539. 33
- Jazwinski, A. H. (1970). Mathematics in science and engineering. *Stochastic processes and filtering theory*, 64. 81
- Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202. 72
- Jongman, B., Ward, P. J., and Aerts, J. C. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, 22(4):823–835. 14
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. 79
- Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. 79
- Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201. 252
- Karamuz, E., Romanowicz, R. J., and Doroszkiewicz, J. (2020). The use of unmanned aerial vehicles in flood hazard assessment. *Journal of Flood Risk Management*, 13(4):e12622. 37, 231
- Kittler, J. and Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition*, 19(1):41–47. 53
- Lai, X., Liang, Q., Yesou, H., and Daillet, S. (2014). Variational assimilation of remotely sensed flood extents using a 2-d flood model. *Hydrology and Earth System Sciences*, 18(11):4325–4339. 16
- Lai, X. and Monnier, J. (2009). Assimilation of spatially distributed water levels into a shallow-water flood model. part i: Mathematical method and test case. *Journal of Hydrology*, 377(1-2):1–11. 36
- Lamboni, M., Monod, H., and Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4):450–459. 72

- Landuyt, L. (2021). *Flood mapping from radar remote sensing using automated image classification techniques*. PhD thesis, Ghent University. 10, 51
- Landuyt, L., Van Wesemael, A., Schumann, G. J.-P., Hostache, R., Verhoest, N. E., and Van Coillie, F. M. (2018). Flood mapping based on synthetic aperture radar: An assessment of established approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):722–739. 54, 125, 126
- Lang, M. W., Townsend, P. A., and Kasischke, E. S. (2008). Influence of incidence angle on detecting flooded forests using c-hh synthetic aperture radar data. *Remote Sensing of Environment*, 112(10):3898–3907. 55
- Le Dimet, F.-X., Castaing, W., Ngnepieba, P., and Vieux, B. (2009). Data assimilation in hydrology: variational approach. *Data assimilation for atmospheric, oceanic and hydrologic applications*, pages 367–405. 74
- Le Dimet, F.-X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):97–110. 85, 91
- Le Ruz, G., Lombardi, D., and Moireau, P. (2025). Optimal filtering in manifolds — a deterministic perspective. *Preprint*. To appear. 233, 247, 250
- Lee, J.-S. (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE transactions on pattern analysis and machine intelligence*, (2):165–168. 46
- Lee, J.-S. and Pottier, E. (2017). *Polarimetric radar imaging: from basics to applications*. CRC press. 46
- Lee, J.-S., Wen, J.-H., Ainsworth, T. L., Chen, K.-S., and Chen, A. J. (2008). Improved sigma filter for speckle filtering of sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):202–213. 46
- Li, Y., Martinis, S., Plank, S., and Ludwig, R. (2018). An automatic change detection approach for rapid flood mapping in sentinel-1 sar data. *International journal of applied earth observation and geoinformation*, 73:123–135. 53
- Lillesand, T., Kiefer, R. W., and Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons. 10, 37, 39, 43, 51
- Limerinos, J. T. (1970). *Determination of the Manning coefficient from measured bed roughness in natural channels*. Number 1898. US Government Printing Office. 33
- Lions, J. L. (1971). *Optimal control of systems governed by partial differential equations*, volume 170. Springer. 91
- Lions, J. L. and Lelong, P. (1968). Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles. 85
- Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. (2022). Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*. 234
- Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528. 91
- L’Ecuyer, P. and Lemieux, C. (2002). Recent advances in randomized quasi-monte carlo methods. *Modeling uncertainty: An examination of stochastic theory, methods, and applications*, pages 419–474. 68

- Manning, R., Griffith, J. P., Pigot, T., and Vernon-Harcourt, L. F. (1890). *On the flow of water in open channels and pipes*. 30
- Marriott, M. and Jayaratne, R. (2010). Hydraulic roughness—links between manning’s coefficient, nikuradse’s equivalent sand roughness and bed grain size. *Advances in Computing and Technology 2010*, pages 27–32. 32
- Martin, M. J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D., Lellouche, J.-M., Mogensen, K., et al. (2015). Status and future of data assimilation in operational oceanography. *Journal of Operational Oceanography*, 8(sup1):s28–s48. 74
- Martinis, S., Kersten, J., and Twele, A. (2015a). A fully automated terrasar-x based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:203–212. 53
- Martinis, S., Kuenzer, C., Wendleder, A., Huth, J., Twele, A., Roth, A., and Dech, S. (2015b). Comparing four operational sar-based water and flood detection approaches. *International Journal of Remote Sensing*, 36(13):3519–3543. 53
- Martinis, S., Plank, S., and Ćwik, K. (2018). The use of sentinel-1 time-series data to improve flood monitoring in arid areas. *Remote Sensing*, 10(4):583. 53, 56
- Martinis, S., Twele, A., and Voigt, S. (2009). Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution terrasar-x data. *Natural Hazards and Earth System Sciences*, 9(2):303–314. 53, 56
- Masafu, C. and Williams, R. (2024). Satellite video remote sensing for flood model validation. *Water Resources Research*, 60(1):e2023WR034545. 232
- Mason, D., Davenport, I., Flather, R., Gurney, C., Robinson, G., and Smith, J. (2001). A sensitivity analysis of the waterline method of constructing a digital elevation model for intertidal areas in ers sar scene of eastern england. *Estuarine, Coastal and Shelf Science*, 53(6):759–778. 59
- Mason, D., Schumann, G.-P., Neal, J., Garcia-Pintado, J., and Bates, P. (2012a). Automatic near real-time selection of flood water levels from high resolution synthetic aperture radar images for assimilation into hydraulic models: A case study. *Remote Sensing of Environment*, 124:705–716. 53
- Mason, D. C., Bates, P., and Dall’Amico, J. (2009a). Calibration of uncertain flood inundation models using remotely sensed water levels. *Journal of Hydrology*, 368(1-4):224–236. 16, 59
- Mason, D. C., Cobby, D. M., Horritt, M. S., and Bates, P. D. (2003). Floodplain friction parameterization in two-dimensional river flood models using vegetation heights derived from airborne scanning laser altimetry. *Hydrological processes*, 17(9):1711–1732. 34
- Mason, D. C., Dance, S. L., Vetra-Carvalho, S., and Cloke, H. L. (2018). Robust algorithm for detecting floodwater in urban areas using synthetic aperture radar images. *Journal of Applied Remote Sensing*, 12(4):045011–045011. 56
- Mason, D. C., Davenport, I. J., Neal, J. C., Schumann, G. J.-P., and Bates, P. D. (2012b). Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images. *IEEE transactions on Geoscience and Remote Sensing*, 50(8):3041–3052. 55
- Mason, D. C., Speck, R., Devereux, B., Schumann, G. J.-P., Neal, J. C., and Bates, P. D. (2009b). Flood detection in urban areas using terrasar-x. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):882–894. 55

- Mateo-Garcia, G., Veitch-Michaelis, J., Smith, L., Oprea, S. V., Schumann, G., Gal, Y., Baydin, A. G., and Backes, D. (2021). Towards global flood mapping onboard low cost satellites with machine learning. *Scientific reports*, 11(1):1–12. 54
- Matgen, P., Hostache, R., Schumann, G., Pfister, L., Hoffmann, L., and Savenije, H. (2011). Towards an automated sar-based flood monitoring system: Lessons learned from two case studies. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7-8):241–252. 53
- Matgen, P., Schumann, G., Henry, J.-B., Hoffmann, L., and Pfister, L. (2007a). Integration of sar-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management. *International Journal of Applied Earth Observation and Geoinformation*, 9(3):247–263. 59
- Matgen, P., Schumann, G., Pappenberger, F., and Pfister, L. (2007b). Sequential assimilation of remotely sensed water stages in flood inundation models. *IAHS PUBLICATION*, 316:78. 36
- Mayo, T., Butler, T., Dawson, C., and Hoteit, I. (2014). Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of manning’s friction coefficient. *Ocean Modelling*, 76:43–58. 33
- McCann, R. J. (1997). A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179. 253
- McKay, M. D., Beckman, R. J., and Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61. 67
- Medeiros, S. C., Hagen, S. C., and Weishampel, J. F. (2012). Comparison of floodplain surface roughness parameters derived from land cover data and field measurements. *Journal of Hydrology*, 452:139–149. 33
- Meliani, M., Bartoli, N., Lefebvre, T., Bouhlef, M.-A., Martins, J. R., and Morlier, J. (2019). Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design. In *AIAA aviation 2019 forum*, page 3236. 67
- Meraoumia, I., Dalsasso, E., Denis, L., Abergel, R., and Tupin, F. (2023). Multitemporal speckle reduction with self-supervised deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14. 47
- Metropolis, N. and Ulam, S. (1949). The monte carlo method journal of the american statistical association, vol. 44, no 247. 67
- Meyer-Peter, E. and Müller, R. (1948). Formulas for bed-load transport. In *IAHSR 2nd meeting, Stockholm, appendix 2*. IAHR. 32
- Miller, R. N., Ghil, M., and Gauthiez, F. (1994). Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of Atmospheric Sciences*, 51(8):1037–1056. 81
- Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704. 250
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174. 69, 70
- Morvan, H., Knight, D., Wright, N., Tang, X., and Crossley, A. (2008). The concept of roughness in fluvial hydraulics and its formulation in 1d, 2d and 3d numerical simulation models. *Journal of Hydraulic Research*, 46(2):191–208. 26, 29, 30, 35

- Mouradi, R.-S. (2021). *Non-linear data-driven modelling on multidimensional fields: an application to hydro-morphodynamic coastal flows*. PhD thesis, Institut National Polytechnique de Toulouse-INPT. 61, 62
- Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R., and Mijic, A. (2020). The use of lidar-derived dem in flood applications: A review. *Remote Sensing*, 12(14):2308. 37
- Neal, J. C., Atkinson, P. M., and Hutton, C. W. (2007). Flood inundation model updating using an ensemble kalman filter and spatially distributed measurements. *Journal of hydrology*, 336(3-4):401–415. 83
- Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., and Horritt, M. S. (2009). Distributed whole city water level measurements from the carlisle 2005 urban flood event and comparison with hydraulic model simulations. *Journal of Hydrology*, 368(1-4):42–55. 36
- Nguyen, N. Y., Ichikawa, Y., and Ishidaira, H. (2016). Estimation of inundation depth using flood extent information and hydrodynamic simulations. *Hydrological Research Letters*, 10(1):39–44. 58, 59
- Nguyen, T. H., Ricci, S., Piacentini, A., Fatras, C., Kettig, P., Blanchet, G., Peña Luque, S., and Baillarin, S. (2022). Dual state-parameter assimilation of sar-derived wet surface ratio for improving fluvial flood reanalysis. *Water Resources Research*, 58(11):e2022WR033155. 233
- Nissen, E., Krishnan, A. K., Arrowsmith, J. R., and Saripalli, S. (2012). Three-dimensional surface displacements and rotations from differencing pre-and post-earthquake lidar point clouds. *Geophysical research letters*, 39(16). 37
- Nodet, M. (2005). *Modélisation mathématique et assimilation de données lagrangiennes pour l’océanographie*. PhD thesis, Université Nice Sophia Antipolis. 61, 62
- Oberstadler, R., Hönsch, H., and Huth, D. (1997). Assessment of the mapping capabilities of ers-1 sar data for flood mapping: a case study in germany. *Hydrological processes*, 11(10):1415–1425. 53, 57, 59
- O’Grady, D., Leblanc, M., and Gillieson, D. (2011). Use of envisat asar global monitoring mode to complement optical data in the mapping of rapid broad-scale flooding in pakistan. *Hydrology and Earth System Sciences*, 15(11):3475–3494. 53
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66. 53
- Panaretos, V. M. and Zemel, Y. (2020). *An invitation to statistics in Wasserstein space*. Springer Nature. 247
- Pappenberger, F., Beven, K., Horritt, M., and Blazkova, S. (2005). Uncertainty in the calibration of effective roughness parameters in hec-ras using inundation and downstream level observations. *Journal of hydrology*, 302(1-4):46–69. 36
- Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances in water resources*, 31(1):1–14. 36
- Pasanisi, A. and Dutfoy, A. (2011). An industrial viewpoint on uncertainty quantification in simulation: Stakes, methods, tools, examples. In *IFIP Working Conference on Uncertainty Quantification*, pages 27–45. Springer. 66
- Paul, F., Barrand, N. E., Baumann, S., Berthier, E., Bolch, T., Casey, K., Frey, H., Joshi, S., Konovalov, V., Le Bris, R., et al. (2013). On the accuracy of glacier outlines derived from remote-sensing data. *Annals of Glaciology*, 54(63):171–182. 37

- Perks, M. T., Russell, A. J., and Large, A. R. (2016). Advances in flash flood monitoring using unmanned aerial vehicles (uavs). *Hydrology and Earth System Sciences*, 20(10):4005–4015. 37
- Peter, B. G., Cohen, S., Lucey, R., Munasinghe, D., Raney, A., and Brakenridge, G. R. (2020). Google earth engine implementation of the floodwater depth estimation tool (fwdet-gee) for rapid and large scale flood analysis. *IEEE Geoscience and Remote Sensing Letters*. 59
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79:214–232. 69, 73
- Pierdicca, N., Pulvirenti, L., Boni, G., Squicciarino, G., and Chini, M. (2017). Mapping flooded vegetation using cosmo-skymed: comparison with polarimetric and optical data over rice fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6):2650–2662. 55
- Powell, M. J. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162. 91
- Puech, C., Hostache, R., Raclot, D., and Matgen, P. (2007). Estimation of flood water levels by merging dem and satellite imagery using hydraulics laws through ai to enhance the estimates. In *Proceedings of Second Space for Hydrology Workshop. ESA. Geneva*, pages 1–7. 58
- Pulvirenti, L., Chini, M., Pierdicca, N., and Boni, G. (2015). Use of sar data for detecting floodwater in urban and agricultural areas: The role of the interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3):1532–1544. 55
- Pulvirenti, L., Marzano, F. S., Pierdicca, N., Mori, S., and Chini, M. (2013). Discrimination of water surfaces, heavy rainfall, and wet snow using cosmo-skymed observations of severe weather events. *IEEE transactions on geoscience and remote sensing*, 52(2):858–869. 55
- Pulvirenti, L., Pierdicca, N., Chini, M., and Guerriero, L. (2011). An algorithm for operational flood mapping from synthetic aperture radar (sar) data using fuzzy logic. *Natural Hazards and Earth System Sciences*, 11(2):529–540. 55, 56
- Raclot, D. (2003). *Méthodologie d’exploitation quantitative des photographies aériennes d’inondation de plaine*. PhD thesis, Université Montpellier II-Sciences et Techniques du Languedoc. 37
- Raclot, D. (2006). Remote sensing of water levels on floodplains: a spatial approach guided by hydraulic functioning. *International Journal of Remote Sensing*, 27(12):2553–2574. 58
- Raclot, D., Maurel, P., and Puech, C. (2003). Apports de la télédétection à très haute résolution spatiale à l’étude du risque d’inondation. *Theses.fr*. 58
- Ramser, C. et al. (1929). Flow of water in drainage channels. *Technical Bulletin*, 179:581–587. 31
- Refice, A., Capolongo, D., Pasquariello, G., D’Addabbo, A., Bovenga, F., Nutricato, R., Lovergine, F. P., and Pietranera, L. (2014). Sar and insar for flood monitoring: Examples with cosmo-skymed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(7):2711–2722. 54, 55
- Ricci, S. (2004). Modélisation multivariée de la matrice de covariance d’erreur d’ébauche pour l’assimilation variationnelle océanique. *These, Université Paul Sabatier-Toulouse III*. 89
- Richardson, L. F. (1922). *Weather prediction by numerical process*. University Press. 74

- Rignot, E. J. and Van Zyl, J. J. (1993). Change detection techniques for ers-1 sar data. *IEEE Transactions on Geoscience and Remote sensing*, 31(4):896–906. 53
- Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J., and Le Dimet, F.-X. (2005). A reduced-order strategy for 4d-var data assimilation. *Journal of Marine Systems*, 57(1-2):70–82. 90
- Robinson, C. (2015). *Image data assimilation with fluid dynamics models: Application to 3D flow reconstruction*. PhD thesis, Université de Rennes. 61, 62
- Rochoux, M. C., Emery, C., Ricci, S., Cuenot, B., and Trouvé, A. (2015). Towards predictive data-driven simulations of wildfire spread—part ii: Ensemble kalman filter for the state estimation of a front-tracking simulator of wildfire spread. *Natural Hazards and Earth System Sciences*, 15(8):1721–1739. 233
- Rochoux, M. C., Ricci, S., Lucor, D., Cuenot, B., and Trouvé, A. (2014). Towards predictive data-driven simulations of wildfire spread—part i: Reduced-cost ensemble kalman filter based on a polynomial chaos surrogate model for parameter estimation. *Natural Hazards and Earth System Sciences*, 14(11):2951–2973. 11, 74, 233
- Rouzies, É. (2023). *Quantification et réduction de l’incertitude dans un modèle de transfert de pesticides à l’échelle du bassin versant*. PhD thesis, Université Grenoble Alpes [2020-....]. 61, 62
- Roy, D. P., Huang, H., Houborg, R., and Martins, V. S. (2021). A global analysis of the temporal availability of planetscope high spatial resolution multi-spectral imagery. *Remote Sensing of Environment*, 264:112586. 231
- Rozet, F. and Louppe, G. (2023a). Score-based data assimilation. *Advances in Neural Information Processing Systems*, 36:40521–40541. 234
- Rozet, F. and Louppe, G. (2023b). Score-based data assimilation for a two-layer quasi-geostrophic model. *arXiv preprint arXiv:2310.01853*. 234
- Salamon, P., Mctormick, N., Reimer, C., Clarke, T., Bauer-Marschallinger, B., Wagner, W., Martinis, S., Chow, C., Böhnke, C., Matgen, P., et al. (2021). The new, systematic global flood monitoring product of the copernicus emergency management service. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 1053–1056. IEEE. 16
- Saltelli, A., Chan, K., Scott, M., et al. (2000). Sensitivity analysis. probability and statistics series. *John and Wiley & Sons, New York*. 69
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons. 69, 70
- Saltelli, A. and Sobol, I. M. (1995). About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering & System Safety*, 50(3):225–239. 71
- Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., et al. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*, volume 1. Wiley Online Library. 68, 70, 71
- Sasaki, Y. (1955). A fundamental study of the numerical prediction based on the variational principle. *Journal of the Meteorological Society of Japan. Ser. II*, 33(6):262–275. 85
- Sasaki, Y. (1958). An objective analysis based on the variational method. *Journal of the Meteorological Society of Japan. Ser. II*, 36(3):77–88. 85

- Savage, J. T. S., Bates, P., Freer, J., Neal, J., and Aronica, G. (2016). When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrological Processes*, 30(13):2014–2032. 14
- Schlaffer, S., Matgen, P., Hollaus, M., and Wagner, W. (2015). Flood detection from multi-temporal sar data using harmonic analysis and change detection. *International Journal of Applied Earth Observation and Geoinformation*, 38:15–24. 53, 54
- Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., and Pappenberger, F. (2009). Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models. *Reviews of Geophysics*, 47(4). 41, 54
- Schumann, G., Cutler, M., Black, A., Matgen, P., Pfister, L., Hoffmann, L., and Pappenberger, F. (2008a). Evaluating uncertain flood inundation predictions with uncertain remotely sensed water stages. *International journal of river basin management*, 6(3):187–199. 58
- Schumann, G., Giustarini, L., Tarpanelli, A., Jarihani, B., and Martinis, S. (2022). Flood modeling and prediction using earth observation data. *Surveys in Geophysics*, pages 1–26. 39, 74, 231
- Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., and Pfister, L. (2007a). High-resolution 3-d flood information from radar imagery for flood hazard management. *IEEE transactions on geoscience and remote sensing*, 45(6):1715–1725. 58, 59
- Schumann, G., Matgen, P., Cutler, M., Black, A., Hoffmann, L., and Pfister, L. (2008b). Comparison of remotely sensed water stages from lidar, topographic contours and srtm. *ISPRS journal of photogrammetry and remote sensing*, 63(3):283–296. 59
- Schumann, G., Matgen, P., Hoffmann, L., Hostache, R., Pappenberger, F., and Pfister, L. (2007b). Deriving distributed roughness values from satellite radar data for flood inundation modelling. *Journal of Hydrology*, 344(1-2):96–111. 36
- Schumann, G., Matgen, P., and Pappenberger, F. (2008c). Conditioning water stages from satellite imagery on uncertain data points. *IEEE Geoscience and Remote Sensing Letters*, 5(4):810–813. 58
- Schumann, G., Pappenberger, F., and Matgen, P. (2008d). Estimating uncertainty associated with water stages from a single sar image. *Advances in water resources*, 31(8):1038–1047. 58
- Schumann, G. J.-P., Neal, J. C., Mason, D. C., and Bates, P. D. (2011). The accuracy of sequential aerial photography and sar data for observing urban flood dynamics, a case study of the uk summer 2007 floods. *Remote Sensing of Environment*, 115(10):2536–2546. 58
- Sellin, R. H., Bryant, T. B., and Loveless, J. H. (2003). An improved method for roughening floodplains on physical river models. *Journal of Hydraulic Research*, 41(1):3–14. 32
- Sezgin, M. and Sankur, B. I. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–168. 53
- Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656. 91
- Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge. 66
- Siripatana, A., Mayo, T., Knio, O., Dawson, C., Le Maître, O., and Hoteit, I. (2018). Ensemble kalman filter inference of spatially-varying manning’sn coefficients in the coastal ocean. *Journal of Hydrology*, 562:664–684. 34

- Smith, L. C. (1997). Satellite remote sensing of river inundation area, stage, and discharge: A review. *Hydrological processes*, 11(10):1427–1439. 41
- Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.*, 1(4):407–414. 71
- Stokes, G. G. (1845). On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids. *Transactions of the Cambridge Philosophical Society*, 8:287–319. 24
- Straatsma, M. and Middelkoop, H. (2006). Airborne laser scanning as a tool for lowland flood-plain vegetation monitoring. *Hydrobiologia*, 565:87–103. 34
- Strickler, A. (1923). Contributions to the question of velocity formula and the roughness numbers for rivers, channels and pipes. 32
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979. 72
- Talagrand, O. and Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. i: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478):1311–1328. 85
- Tanguy, M., Chokmani, K., Bernier, M., Poulin, J., and Raymond, S. (2017). River flood mapping in urban areas combining radarsat-2 data and flood return period data. *Remote Sensing of Environment*, 198:442–459. 56
- Tarpanelli, A., Brocca, L., Melone, F., and Moramarco, T. (2013). Hydraulic modelling calibration in small rivers by using coarse resolution synthetic aperture radar imagery. 40
- Tarpanelli, A., Mondini, A. C., and Camici, S. (2022). Effectiveness of sentinel-1 and sentinel-2 for flood detection assessment in europe. *Natural Hazards and Earth System Sciences*, 22(8):2473–2489. 42
- Thanh Noi, P. and Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. *Sensors*, 18(1):18. 54
- Thual, O. (2010). *Hydrodynamique de l’environnement*. Ecole polytechnique Palaiseau. 23
- Tikhonov, A. N. and Arsenin, V. (1977). Solutions of ill-posed problems. 76
- Travert, J.-P., Goeury, C., Bacchi, V., Zaoui, F., and Boyaval, S. (2024). Influence of 3d-var dissimilarity measure on flood model calibration with satellite data. *XXXth TELEMAC User Conference*. 206
- Tsyganskaya, V., Martinis, S., Marzahn, P., and Ludwig, R. (2018). Sar-based detection of flooded vegetation—a review of characteristics and approaches. *International journal of remote sensing*, 39(8):2255–2293. 55
- Twele, A., Cao, W., Plank, S., and Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing*, 37(13):2990–3004. 53, 56
- Ulaby, F. T. (1982). Radar remote sensing and surface scattering and emission theory. *Microwave remote sensing: active and passive*. 41, 52
- Ulaby, F. T., Michielssen, E., and Ravaioli, U. (2015). *Fundamentals of applied electromagnetics*, volume 7. Pearson Upper Saddle River, NJ. 45, 47

- Ulaby, F. T., Moore, R. K., and Fung, A. K. (1986). Microwave remote sensing: Active and passive. volume 3-from theory to applications. 44
- Van der Sande, C., De Jong, S., and De Roo, A. (2003). A segmentation and classification approach of ikonos-2 imagery for land cover mapping to assist flood risk and flood damage assessment. *International Journal of applied earth observation and geoinformation*, 4(3):217–229. 34
- Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114. 85
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media. 54
- Voigt, S., Giulio-Tonolo, F., Lyons, J., Kučera, J., Jones, B., Schneiderhan, T., Platzeck, G., Kaku, K., Hazarika, M. K., Czarán, L., et al. (2016). Global trends in satellite-based emergency mapping. *Science*, 353(6296):247–252. 16
- Weaver, A. and Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127(575):1815–1846. 89
- Weaver, A., Vialard, J., and Anderson, D. (2003). Three-and four-dimensional variational assimilation with a general circulation model of the tropical pacific ocean. part i: Formulation, internal diagnostics, and consistency checks. *Monthly Weather Review*, 131(7):1360–1378. 92
- Werner, M., Hunter, N., and Bates, P. (2005). Identifiability of distributed floodplain roughness values in flood extent estimation. *Journal of Hydrology*, 314(1-4):139–157. 36
- Westerhoff, R., Kleuskens, M., Winsemius, H., Huizinga, H., Brakenridge, G., and Bishop, C. (2013). Automated global water mapping based on wide-swath orbital synthetic-aperture radar. *Hydrology and Earth System Sciences*, 17(2):651–663. 54
- Wieland, M. and Martinis, S. (2019). A modular processing chain for automated flood monitoring from multi-spectral satellite data. *Remote Sensing*, 11(19):2330. 54
- Wohl, E. E. (1998). Uncertainty in flood estimates associated with roughness coefficient. *Journal of Hydraulic Engineering*, 124(2):219–223. 13, 33
- Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., Corato, G., Matgen, P., and Bates, P. (2016). Calibration of channel depth and friction parameters in the lisflood-fp hydraulic model using medium-resolution sar data and identifiability techniques. *Hydrology and Earth System Sciences*, 20(12):4983–4997. 16, 37
- Woodhouse, I. H. (2017). *Introduction to microwave remote sensing*. CRC press. 45, 47
- Wooster, M. J., Roberts, G., Smith, A. M., Johnston, J., Freeborn, P., Amici, S., and Hudak, A. T. (2013). Thermal remote sensing of active vegetation fires and biomass burning events. In *Thermal infrared remote sensing: sensors, methods, applications*, pages 347–390. Springer. 37
- Yen, J.-C., Chang, F.-J., and Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing*, 4(3):370–378. 53
- Zhao, B., Sui, H., Liu, J., Shi, W., Wang, W., Xu, C., and Wang, J. (2024). Flood inundation monitoring using multi-source satellite imagery: A knowledge transfer strategy for heterogeneous image change detection. *Remote Sensing of Environment*, 314:114373. 231

- Zhao, J., Pelich, R., Hostache, R., Matgen, P., Cao, S., Wagner, W., and Chini, M. (2021). Deriving exclusion maps from c-band sar time-series in support of floodwater mapping. *Remote Sensing of Environment*, 265:112668. [53](#), [56](#)
- Zhongming, Z., Linong, L., Xiaona, Y., Wangqiang, Z., Wei, L., et al. (2021). Ar6 climate change 2021: The physical science basis. [14](#)
- Ziliani, M. G., Ghostine, R., Ait-El-Fquih, B., McCabe, M. F., and Hoteit, I. (2019). Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation. *Journal of Hydrology*, 577:123924. [83](#)
- Zwenzner, H. and Voigt, S. (2009). Improved estimation of flood parameters by combining space based sar data with very high resolution digital elevation data. *Hydrology and Earth System Sciences*, 13(5):567–576. [58](#)